REMARKS

Claims 1-3 and 5-9 are currently pending and under examination in the application. Claims 11 and 12 were previously withdrawn as being directed to non-elected subject matter.

Claims 1 and 5 have been amended solely to expedite patent prosecution in accordance with the U.S. Patent Office Business Goals (65 Fed. Reg. 54604 (September 8, 2000)). Applicant reserves the right to present any cancelled subject matter in a co-pending application.

Amended claims 1 and 5 recite "protein-protein interaction" and "where the plurality of proteins are screened concurrently" (see, *inter alia*, page 1, lines 12-15; page 4, lines 10-12; page 7, lines 9-12; page 9, lines 9-22 to page 10, lines 6-22; page 16, lines 3-10; page 18, lines 10-16; page 19, lines 1-20 to page 20, lines 1-8; and Example I, page 22, lines 1-19; and Examples II-IV, page 23, lines 3-22 of the application as originally filed).

Newly added claim 13 recites "[a] method of correlating proteomic interaction(s) with oxygen tension comprising (a) screening for protein levels of a plurality of proteins, where the screening is performed in room air, and where the plurality of proteins are screened concurrently; (b) screening for protein levels of a plurality of proteins, where the screening is performed in the presence of decreased oxygen tension, and where the plurality of proteins are screened concurrently; and (c) correlating the protein level(s) with oxygen tension by identifying at least one different protein level between (a) and (b)" (see, *inter alia*, page 7, lines 6-12; and Example IV, page 23, lines 18-22 of the originally filed application)..

Newly added claim 14 recites "[t]he method of Claim 13 where at least one protein employed in the determination is associated with a physiological process or a pathophysiological process" (see, *inter alia*, page 7, lines 6-12 of the application as originally filed).

Newly added claim 15 recites "[t]he method of Claim 13 where a plurality of determinations are made in step (b) with different oxygen tensions being employed in each determination" (see, *inter alia*, page 20, lines 1-3 of the application as originally filed).

Newly added claim 16 recites "[t]he method of Claim 15 where the oxygen tensions

employed are in step (b) range from 0.1 mm Hg to 145 mm Hg" (see, *inter alia*, page 19, lines 19-20 of the application as originally filed).

Newly added claim 17 recites "[t]he method of Claim 13 where the different protein levels in step (c) are used to identify protein functions associated with a pathophysiological process" (see, *inter alia*, page 20, lines 6-8 of the application as originally filed).

These amendments are supported by the application as originally filed, and do not constitute new matter. Specific support for the amendments is shown in parentheses, above. Entry of these amendments in the application is respectfully requested.

Previous Rejections

The Examiner has withdrawn all previous rejections from the Office Action mailed July 13, 2004 (Office Action, page 2).

35 U.S.C. §112, First Paragraph

Claims 1-3 and 5-9 have been rejected under 35 U.S.C. §112, first paragraph, as allegedly failing to comply with the written description requirement (Office Action, page 3). The Examiner states that the claims contain subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventor has possession of the claimed invention at the time the application was filed. *Id.* The Examiner states that the phrase "a proteomic interaction between at least one protein and a plurality of proteins" includes new matter. *Id.*

Applicant respectfully disagrees.

The application specifically teaches that proteomic interactions (e.g., protein-protein interactions, protein-DNA interactions, protein activity, and protein expression levels) can be screened by several well-known means (see, *inter alia*, page 9, lines 3-22 to page 11, lines 1-8 of the originally filed application). However, solely for the purpose of expediting patent prosecution, claims 1 and 5 have been amended and claim 13 has been added as follows.

1. A method of establishing a protein-protein interaction map comprising
(a) screening for a protein-protein interaction between at least one protein and a plurality of proteins, where the screening is performed in the absence of a simulated redox state perturbation, and where the plurality of proteins are screened concurrently;
(b) screening for a protein-protein interaction between the at least one protein and a plurality of proteins, where the screening is performed in the presence of a simulated redox perturbation, and where the plurality of proteins are screened concurrently; and
(c) generating the protein-protein interaction map by identifying at least one different protein-protein interaction between (a) and (b).

5. A method of correlating protein-protein interaction(s) with oxygen tension comprising
(a) screening for a protein-protein interaction between at least one protein and a plurality of proteins, where the screening is performed in room air, and where the plurality of proteins are screened concurrently;
(b) screening for a protein-protein interaction between the at least one protein and a plurality of proteins, where the screening is performed in the presence of decreased oxygen tension, and where the plurality of proteins are screened concurrently; and
(c) correlating the protein-protein interaction(s) with oxygen tension by identifying at least one different protein-protein interaction between (a) and (b).

13. A method of correlating protein levels(s) with oxygen tension comprising
(a) screening for protein levels of a plurality of proteins, where the screening is performed in room air, and where the plurality of proteins are screened concurrently;
(b) screening for protein levels of a plurality of proteins, where the screening is performed in the presence of decreased oxygen tension, and where the plurality of proteins are screened concurrently; and
(c) correlating the protein level(s) with oxygen tension by identifying at least one different protein level between (a) and (b).

The Examiner states that the instant specification provides support only for protein-protein interactions measured by two-hybrid systems (Office Action, page 4). Yet, the application specifically teaches that protein-protein interactions and protein levels can be screened by several well-established techniques, including those published by:

- Fung et al., 2001 (protein chips, microfluidic protein chips, peptide arrays, surface plasmon resonance (SPR), surface-enhanced laser desorption/ionization (SELDI), and time of flight-mass spectroscopy (TOF-MS); Exhibit 1, see page 65, right column to page 66, left column; and page 67, left column),

- Delneri et al., 2001 (two-dimensional gel electrophoresis, multidimensional protein identification technique (MudPIT), matrix-assisted laser desorption/ionization mass spectroscopy (MALDI-MS), and isotope-coded affinity tagging (ICAT); Exhibit 2, see page 88, right column to page 89, left column),

- Zhu et al., 2001 (proteome chips; Exhibit 3, see page 2101, right column to page 2102, left and center columns), and

- Sakurai et al., 1998 (orphan receptor cell lines; Exhibit 4, see page 573, right column, bottom) (see, *inter alia*, page 9, lines 3-22 to page 11, lines 1-8 of the originally filed application).

For the Written Description requirement, it is noted that Applicant need not disclose in detail those aspects of the invention which are well known to persons of ordinary skill in the art. *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1379-80, and 1384 (Fed. Cir. 1986); MPEP § 2163. This is particularly true for mature technologies, where publications and patents are available to skilled artisans. *See In re Hayes Microcomputer Products, Inc.*, 982 F.2d 1527, 1534-35 (Fed. Cir. 1992); MPEP §2163; *see*, also, Exhibits 1-4.

Moreover, it is not necessary for the application to describe *exactly* the claimed subject matter, but only to convey with *reasonably clarity* that Applicant was in possession thereof. *Union Oil Co. v. Atlantic Richfield Co.*, 208 F.3d 989, 997 (Fed. Cir. 2000). Here, Applicant's claims encompass methods of screening protein-protein interactions and protein levels and the instant application describes numerous well known and widely published approaches for performing such screens.

For at least the reasons set forth above, claims 1-3 and 5-9 presented herein, as well as newly added claims 13-17, are supported by the application as originally filed. Reconsideration is respectfully requested.

35 U.S.C. §102(b)

Claims 1-3 have been rejected under 35 U.S.C. 102(b) as allegedly being anticipated by Cominacini et al., 1997, *Free Radical Biology & Medicine*, 22:117-127 ("Cominacini"; Office Action, page 4). According to the Office Action, Cominacini reports the expression levels of ICAM-1, VCAM-1, and E-selectin in the presence and absence of oxidized LDL protein (Office Action, page 5).

Applicant respectfully traverses.

Anticipation under 35 U.S.C. §102 requires that the cited reference teach every aspect of the claimed invention. *Verdegaal Bros. v. Union Oil Co.*, 814 F.2d 628, 631 (Fed. Cir. 1987); MPEP §706.02 IV. Thus, "[t]he identical invention must be shown in as complete detail as contained in the...claim." *Richardson v. Suzuki Motor Co.*, 868 F.2d 1226, 1236 (Fed. Cir. 1989); MPEP §2131.

Here, the claimed methods include i) screening for a protein-protein interaction between at least one protein and a plurality of proteins; ii) screening for protein-protein interaction(s) in room air and reduced oxygen tension; or iii) screening for protein levels of a plurality of proteins in room air and in reduced oxygen tension, none of which are taught or suggested by Cominacini.

For at least these reasons, Cominacini cannot anticipate the subject matter of claims 1-3 or 5-9 presented herein, or newly added claims 13-17. Reconsideration is respectfully requested.

35 U.S.C. §103(a)

Claims 1-3 have been rejected under 35 U.S.C. 103(a) as allegedly being unpatentable in view of Nishiyama et al., 1999, *J. Biol. Chem.*, 274:21645-50 ("Nishiyama"; Office Action, page 6). According to the Office Action, Nishiyama reports the binding interaction between TRX and

TBP-2 in the presence of redox modifiers. *Id.* The Examiner states that it would have been obvious to one of skill in the art to characterize the binding interaction of TRX with other candidates. *Id.* The Examiner states that this characterization would merely involve the duplication of steps used for TRX and TBP. *Id.*

Applicant respectfully traverses.

As amended, claim 1 reads:

> 1. A method of establishing a protein-protein interaction map comprising
> (a) screening for a protein-protein interaction between at least *one protein and a plurality of proteins*, where the screening is performed in the absence of a simulated redox state perturbation, and *where the plurality of proteins are screened concurrently*;
> (b) screening for a protein-protein interaction between the at least *one protein and a plurality of proteins*, where the screening is performed in the presence of a simulated redox perturbation, and *where the plurality of proteins are screened concurrently*; and
> (c) generating the protein-protein interaction map by identifying at least one different protein-protein interaction between (a) and (b). (Emphasis added).

For analysis under 35 U.S.C. §103, it is essential to consider all of the elements of the claimed invention. *W.L. Gore & Associates, Inc. v. Garlock, Inc.*, 721 F.2d 1540, 1548 (Fed. Cir. 1983); *Jones v. Hardy*, 727 F.2d 1524, 1530 (Fed. Cir. 1984); MPEP §2141.02. Each express claim limitation must be taken into account. *See, e.g., Bausch & Lomb v. Barnes-Hind/Hydrocurve, Inc.*, 796 F.2d 443, 447-49 (Fed. Cir. 1986); MPEP §2141.02.

Here, Applicant claims encompass a method of screening for a protein-protein interaction between at least *one protein and a plurality of proteins*, where the screening is performed in the *presence and absence of a simulated redox state perturbation* and *where the plurality of proteins are screened concurrently.*

Yet, Nishiyama does not teach or suggest at least these elements in the pending claims. Nishiyama reports only the interaction of *one* protein (TRX) with one other protein (TBP-2) (see, e.g., Figure 3). Nishiyama does not teach or suggest the screening of at least one protein and a

plurality of proteins. Nor does Nishiyama teach or suggest the concurrent screening of the plurality of proteins. As such, Nishiyama cannot make obvious the subject matter of the present claims.

According to the Office Action, Nishiyama shows that the interaction between TRX and TBP-2 is inhibited by the presence of certain redox reagents (Office Action, page 6). However, Nishiyama reports that TRX was treated with "reducing/oxidizing reagents for 15 min at room temperature and then *washed [five times] with degassed...buffer* and subjected to a binding assay..." (Nishiyama, see page 21646, right column, and Figure 3).

Thus, the reducing/oxidizing reagents are *removed* by Nishiyama by several rounds of washing that precede the binding assay. In contrast, Applicant teaches and claims screens where the screening is performed in the *presence and absence of a simulated redox state perturbation* (see above). Because Nishiyama does not teach or suggest at least these elements in the pending claims, it cannot make obvious the subject matter of these claims.

Moreover, although a claimed invention is alleged to be within the capabilities of one of ordinary skill in the art, this is not sufficient by itself to establish obviousness. MPEP §2143.01; *In re Kotzab*, 217 F.3d 1365, 1371 (Fed. Cir. 2000). Even where a prior art reference may be capable of being modified as claimed, there must be a suggestion or motivation in the reference to do so in order to establish obviousness. *In re Mills*, 916 F.2d 680, 682 (Fed. Cir. 1990); MPEP §2143.01.

Here, there is no suggestion or motivation to modify Nishiyama to produce method of screening for a protein-protein interaction between at least *one protein and a plurality of proteins* and *where the plurality of proteins are screened concurrently*. Even assuming *arguendo* that the steps of Nishiyama could be duplicated for other TRX binding candidates (which Applicant contests), this would involve the process of:

i) Cloning the binding candidate from the cDNA library; ii) Subcloning the coding region from the binding candidate into a TNT expression vector; iii) Preparing and isolating the TRX-GST fusion protein; iv) Treating the TRX-GST fusion protein with DTT, diamide, or hydrogen

peroxide; v) Preparing and isolating *in vitro* translated, ^{35}S-labeled binding candidate; vi) Washing the TRX-GST fusion protein; vii) Incubating the TRX-GST fusion protein with the ^{35}S-labeled binding candidate; and vii) Analyzing the binding by SDS-PAGE analysis and autoradiography.

As such, Nishiyama is not amenable to concurrent (e.g., high throughput) screening and the "mere duplication" of steps in Nishiyama cannot be used to obtain the claimed invention.

Applicant notes that the motivation to modify references under 35 U.S.C. §103 must come from the prior art, itself. *In re Mills*, 916 F.2d at 682. Thus, it is essential:

> "to forget what...has been taught about the claimed invention and cast the mind back to the time the invention was made to occupy the mind of one skilled in the art who is presented only with the references and...the then-accepted wisdom in the art." *In re Fine*, 837 F.2d at 1075 (citing *W.L. Gore*, 721 F.2d at 1553); MPEP §2141.01.

It is impermissible to use the claimed invention as an instruction manual or "template" to piece together the teachings of the cited references so that the claimed invention is alleged to be obvious. *In re Fritch*, 972 F.2d at 1266.

In the instant case, Nishiyama does not teach or suggest all of the elements in the pending claims and there is no independent suggestion or motivation to modify or combine the reports of Nishiyama to obtain the claimed invention. For at least these reasons, pending claims 1-3 cannot be considered unpatentable over Nishiyama and a *prima facie* case of obviousness has not been established. Withdrawal of this rejection is respectfully requested.

CONCLUSION

A favorable action on the merits is respectfully requested. If further discussion of this case is deemed helpful, the Examiner is encouraged to contact the undersigned at the telephone number provided below, and is assured of full cooperation in progressing the instant claims to allowance.

Applicant believes no further fee is due at this time; however, the Commissioner is authorized to charge any additional fees that may be due, or to credit any overpayment, to the undersigned's account, Deposit Account No. 50-0311, Reference No. 24862-503, Customer No. 35437.

Date: July 18, 2005

Respectfully submitted,

Ivor R. Elrifi, Reg. No. 39,529
Caryn DeHoratius, Reg. No. 45,881
Attorneys for Applicant
MINTZ, LEVIN, COHN, FERRIS,
GLOVSKY and POPEO, P.C.
666 Third Avenue, 24th Floor
New York, New York 10017
Telephone: (212) 935-3000
Telefax: (212) 983-3115

NYC 281168v1

Protein biochips for differential profiling

Eric T Fung*, Vanitha Thulasiraman†, Scot R Weinberger‡ and Enrique A Dalmasso§

Progress has been made in utilizing ProteinChip® technology to profile and compare protein expression in normal and diseased states, particularly in the areas of cancer, infectious disease and toxicology. The past year has also seen the development of several novel chip types designed to analyze proteins in a fashion analogous to the array-based format of DNA microarrays. Some of these platforms may be used for differential profiling.

Addresses
Ciphergen Biosystems, 6611 Dumbarton Circle, Fremont, CA 94555, USA
*e-mail: efung@ciphergen.com
†e-mail: vanitha@ciphergen.com
‡e-mail: sweinberger@ciphergen.com
§e-mail: edalmasso@ciphergen.com

Current Opinion in Biotechnology 2001, 12:65–69

0958-1669/01/$ – see front matter

Abbreviations

ACPm	acyl carrier protein
INH	isoniazide
LCM	laser capture microdissection
MALDI	matrix-assisted laser desorption/ionization
PSA	prostate-specific antigen
PSMA	prostate-specific membrane antigen
SELDI	surface-enhanced laser desorption/ionization
TOF	time-of-flight

Introduction

Virtually all strategies for studying cellular function require the comparison of control states with perturbed states. At the heart of all such strategies is the requirement for an assay that sensitively and specifically measures differences in gene or protein expression or function. In the genomic and proteomic era, these assays no longer limit themselves to individual or small sets of genes or proteins; rather, they seek to examine the full complement of protein expression in a cell under a given set of physiological conditions. Because this represents a daunting task, protein differential display techniques that compare protein profiles between control and experimental populations have become increasingly useful. In these systems, expression of proteins common to both groups is ignored and emphasis is placed upon identifying and quantifying those proteins whose expression level is either upregulated or downregulated. These proteins become potential biomarkers representative of a given metabolic disease, drug reaction, neoplasm or microbial infection and are thus diagnostic or possible therapeutic targets.

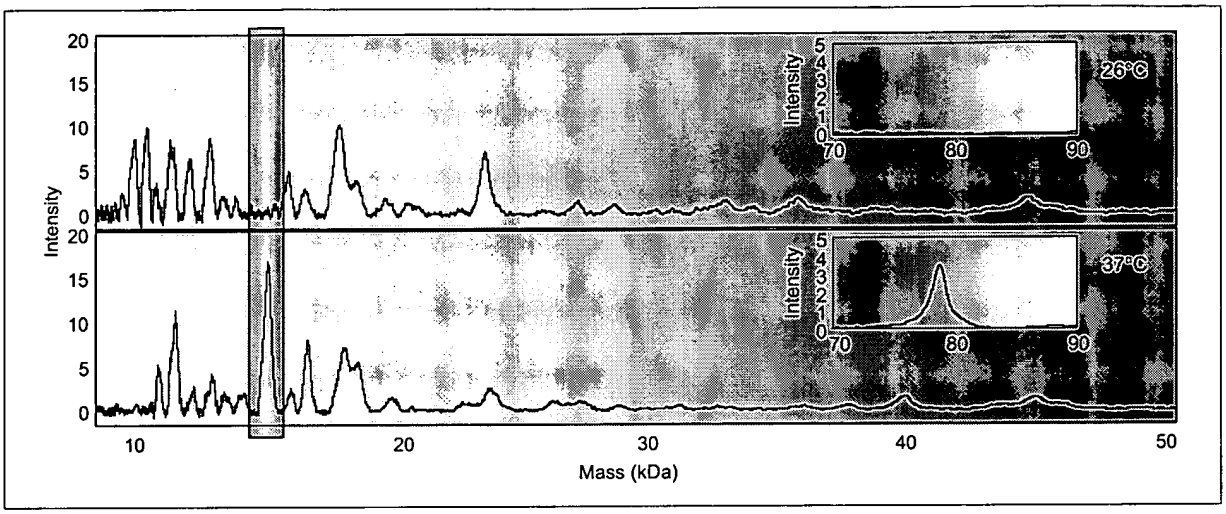
Chip-based protein differential display systems have, for several reasons, been significantly more difficult to develop than their counterparts that examine differential gene expression. First, whereas the reverse-transcriptase polymerase chain reaction (RT-PCR) permits the amplification of mRNA, there is no analogous method to amplify protein expression. Second, for most functional assays (e.g. interaction studies), the proteins must be immobilized on the surface of the chip such that they retain their native conformation and also such that their active site(s) are exposed rather than buried. Additionally, the heterogeneity of the biophysical properties of proteins makes it virtually impossible for any single surface chemistry to bind the full complement of proteins present in a cell, particularly if one attempts to maintain the proteins in their active state. Finally, the capacity of the chip must be sufficient to allow as complete a representation of the proteome to be visualized as possible; abundant proteins (metabolic and cytoskeletal) overwhelm the detection of less abundant proteins such as signaling molecules and receptors, which are generally of more therapeutic interest. Strategies used to circumvent these problems are discussed by Weinberger *et al.* [1].

In this review, we focus on a subset of chip-based assays that may be used to compare protein expression in normal and perturbed states. The most advanced commercially available system for this purpose is the ProteinChip® proteomics platform, and so most of the review will be devoted to describing applications of this system. The review will also describe briefly some of the more novel and unique chip-based assays that have been described in the literature. Because of space limitations, we cannot cover all chip-based proteomics formats. Another commonly used chip technique, surface plasmon resonance (SPR), is described in greater detail by Scheller *et al.* in this issue (pp 35–40) and by Rich and Myszka [2] in a recent issue of this journal. We refer the reader to reviews by Figeys [3] and Guetens *et al.* [4] for discussions of microfluidic protein biochips, and peptide arrays are reviewed by Schneider-Mergener and colleagues in this issue (pp 59–64).

ProteinChip® proteomics

Perhaps the most established chip-based proteomics profiling platform is the incorporation of ProteinChip® technology with mass spectrometry, as commercially available from Ciphergen Biosystems, Inc. (reviewed in [5•]). At the heart of this technology are the ProteinChip® arrays, which have varying chromatographic properties, for example anion exchange, cation exchange, metal affinity and reverse phase. A complex mixture of proteins, as from cells or body fluids, can be reduced to sets of proteins with common properties by binding the sample to chips with differing surface chemistries in parallel and in series. After the chips are washed to remove unbound proteins, the bound proteins are

Figure 1



Protein profiling of *Yersinia pestis* cultured at its two physiological temperatures. 10 µg of crude cytosolic extracts of *Y. pestis* grown at 26°C versus 37°C were analyzed on a strong anion exchange chip (SAX-2). Proteins expressed only at 37°C, selected for purification and identification, were the 14.9 kDa (boxed) and 78.8 kDa (inset) proteins. The 14.9 kDa protein was identified as antigen 4 and the 78.8 kDa protein as the catalase/peroxidase KatY protein. Adapted from [6] with permission.

read in a time-of-flight mass spectrometer (TOF MS). The resulting spectra give a multidimensional binding picture on the basis of different types of interaction. This process, known as surface-enhanced laser desorption/ionization (SELDI), has several advantages over matrix-assisted laser desorption/ionization (MALDI). As in MALDI, the sample is admixed with a small acidic molecule (the matrix) that crystallizes around the sample. In SELDI, however, the ProteinChip® array acts as a surface to which the sample binds uniformly and the matrix is placed on the chip only after the proteins are bound to the chip. Consequently, the spectra obtained are more uniform and reproducible as compared with MALDI-obtained spectra. This difference enables relative protein quantitation with SELDI that is not possible with MALDI. Although SELDI has numerous applications, we focus in this review on one specific application, protein profiling, and refer the reader to the review by Merchant and Weinberger [5•] for descriptions of some of its other applications.

Protein profiling or protein differential display studies have been widely used in the area of disease research, as comparison of lysates from normal versus diseased cells can reveal the expression of important marker proteins. Lysates from disease and control samples are processed on the same types of chip surfaces and the chips are read under the same data collection conditions. Subsequent peak comparison allows the identification of differences. Once a peak of interest has been detected, the analyte can be enriched or purified for further analysis. This is possible by washing the chip with varying stringencies of pH, salt or organic solvent, depending on the type of chip surface chemistry. Once a peak is sufficiently purified, on-chip digestion with proteolytic enzymes followed by analysis of the peptide patterns can yield important identification information. At the current stage of the technology, protein identification applications are still somewhat limited by the mass accuracy of the ProteinChip® reader and it is therefore important to use careful calibration techniques with known peptides of comparable molecular weight to obtain reliable peptide maps. Even with a mass accuracy in the milliDalton range, the proteomic researcher is cautioned that the protein chest of nature contains enough redundancy and sequence conservation to make wrong identification calls on the basis of peptide masses. The combination of profiling using ProteinChip® arrays with powerful fragmentation and sequencing capabilities can enable the researcher to obtain high-resolution data and novel protein information directly from biological samples.

As an example of profiling followed by protein identification in the realm of infectious disease, the ProteinChip® technology has been used to compare the pattern of protein expression in two physiological states of *Yersinia pestis*, the causative agent of the plague. Because it exists in two carriers, the flea and the rodent, *Y. pestis* has evolved to express different sets of proteins at the different temperatures encountered in the carrier organisms (26°C and 37°C). Thulasiraman *et al.* [6] have used ProteinChip® technology to identify two proteins whose expression is regulated by temperature. Lysates of *Y. pestis* grown at the two different temperatures were fractionated on strong anion exchange and immobilized metal-affinity chromatography chips, and the protein expression profiles were

compared (Figure 1). Two of the proteins seen to be upregulated at 37°C were purified and identified as the catalase/peroxidase KatY protein (78.8 kDa) and antigen 4 (14.9 kDa) via mass spectrometric analysis of tryptic peptide fingerprints.

ProteinChip® technology is also useful in linking gene-array expression data with protein discovery and, unlike its gene-based counterpart, can be used to examine post-translational modification of proteins. This approach was applied to verify upregulation of the acyl carrier protein (ACPm) in *Mycobacterium tuberculosis* treated with isoniazide (INH), an antifungal agent known to disrupt both micolic acid and cell wall synthesis. Transcription-array studies of INH-treated *M. tuberculosis* indicated a 4.3-fold increase in ACPm mRNA levels after a six hour exposure to INH. Lysates of control and INH-treated *M. tuberculosis* cultures were examined using anion exchange ProteinChip® arrays and SELDI-TOF MS, followed by analysis of an on-chip protein digest of a 13,217 Da protein using a laser desorption/ionization triple quadrupole TOF MS capable of MS/MS operation [7]. This protein, which was determined to be ACPm conjugated with its cofactor (phosphopantetheine) and a C_{26} fatty acid, appeared to be downregulated on the basis of ProteinChip® array data. However, closer examination of the ProteinChip® array data revealed that peaks found at 12,943.9, 12,997.7 and 13,025.0 Da in the INH-exposed lysates were in fact ACPm–phosphopantetheine conjugated with C_6, C_{10} and C_{12} fatty acids, respectively. In this manner, the presence of INH disrupted proper cell wall synthesis, reducing the abundance of the ACPm–C_{26} conjugate while elevating levels of the smaller fatty acid varieties to a point where total ACPm abundance was upregulated when compared with the control group.

Cancer research in particular has embraced platforms that enable differential protein profiling, as the identification of upregulated or downregulated proteins suggests the presence of a tumor marker that might be used for diagnosis, prognosis monitoring of disease progression or therapeutic success and/or as a therapeutic target. Cancer specimens can be compared with normal specimens, or cancerous regions of a biopsy or specimen can be directly compared with non-cancerous regions of the same specimen. Specific regions of a specimen can be selected for analysis using a technique called laser capture microdissection (LCM) [8]. Petricoin and colleagues [9] have used LCM in conjunction with ProteinChip® proteomics to study protein expression profiles in patient-matched normal colon, cancerous and metastatic samples with the generation of distinct protein profiles for each group. In particular, the authors used this approach to study normal prostate, prostatic intraepithelial neoplasia (widely considered to be a precursor lesion to invasive carcinoma) and prostate cancer, and they showed that the cancer specimens demonstrated downregulation of a 28 kDa protein. A similar study was performed on head and neck cancers [10].

Although SELDI studies are routinely performed on tissue samples, isolation of markers from body fluids, particularly serum or urine, may prove more valuable for diagnostic purposes. Wright *et al.* [11] have taken the approach of studying both tissue and body fluids in the search for biomarkers of prostate cancer. In this study, the authors showed that ProteinChip® proteomics could be used to identify known prostate cancer-associated biomarkers (prostate-specific antigen [PSA], prostatic acid phosphatase, prostate-specific membrane antigen [PMSA] and prostate-specific peptide) in cell lysates of LCM-captured prostatic cells and body fluids from prostate cancer patients, as well as to discover several potential biomarkers, including a 33 kDa and an 18 kDa protein found to be upregulated in prostate cancer cells. When the authors pre-fractionated the samples by size exclusion chromatography followed by ionic exchange chromatography before binding to the ProteinChip® array, they could detect over 300 protein peaks. On the basis of this approach, over 30 proteins were detected as being either over-expressed or under-expressed in specimens from prostate cancer patients [12]. Although no single protein was found to distinguish prostate cancer from the non-cancer groups, a combination of multiple protein peaks was capable of discriminating prostate cancer from a normal age-matched healthy male population. These studies were performed on a variety of body fluids, including serum, urine and seminal plasma.

The ProteinChip® platform used to identify biomarkers can also be used to develop a rapid, sensitive and high-throughput multi-marker assay. The premise of this approach is first to establish composite fingerprint profiles of both disease and non-disease states from a series of training samples, and then to use these composite profiles to make a diagnosis on actual unknown patient samples. In this case it is not essential to identify the proteins to make a diagnosis. Moreover, by utilizing a group of biomarkers one is not constrained by the sensitivity and specificity of any single biomarker, which may be relatively low. For example, by evaluating multiple urinary proteins by SELDI, the detection rate for low stage/grade bladder cancer increased to greater than 75%, compared with 30% by conventional urine cytology [13]. Another potential clinical application of the SELDI system is to develop immunoassays by immobilizing antibody to a specific biomarker on the ProteinChip® array surface. Wright and co-workers [14,15] have successfully used this approach to measure free PSA and complexed PSA, and PSMA in serum and seminal plasma. Interestingly, this platform was successful for quantitation of PSMA in serum where other immunoassay formats had failed. Also of interest was the detection of other possible PSMA isoforms that appear to be differentially expressed in normal, benign and malignant prostate samples, and that would not have been detected with other assay formats. By adding a variety of antibodies to a single array, it is possible to develop a multiplex assay to measure simultaneously eight or more biomarkers using a single ProteinChip® array.

New chip platforms for protein expression analysis

In recent months, several important papers have described the development of novel chip techniques, all of which employ the same concept as DNA arrays — to place thousands of addressable spots in a small grid. Emili and Cagney [16] propose segregating such arrays into nonliving and living arrays. Nonliving arrays consist of grids of proteins or peptides and are assayed biochemically, whereas living arrays consist of grids of organisms, typically bacteria or yeast, and are assayed metabolically. The first large-scale array of proteins was developed as a 'living array' to study two-hybrid interactions — yeast colonies expressing specific combinations of proteins were spotted in known positions and their phenotype (growth on selective media) was assayed [17]. Such a platform could be adapted for profiling analysis by constructing numerous types of selective media and replica plating grids of yeast onto each substrate. Extension of such a principle to the study of mammalian cells (e.g. tumor cell lines) will prove more technically challenging.

Another example of a living array is the antibody array recently described by de Wildt *et al.* [18••]. The arrays created by these authors contain a collection of 18,342 bacterial clones, each expressing a different single-chain antibody. These are spotted onto a 22 × 22 cm filter in duplicate, or duplicate filters might be used. It may be possible to use this technique in the future to randomly generate antibodies, to create arrays of these randomly generated antibodies, to incubate lysates of relevant tissue or even body fluids with the filters and to identify upregulated or downregulated proteins. The proteins of interest can then be enriched and identified using, for example, MALDI-MS.

Recently, MacBeath and Schreiber [19•] have described a novel protein microarray that utilizes glass slides coated with an aldehyde-containing silane reagent. The aldehydes react with primary amines (contained primarily in lysines and at the amino terminus) exposed on the protein's surface, leading to the binding of proteins in many different conformations on the chip. In a small-scale test of this chip, protein–protein interaction studies were performed — baits were printed onto the slide and probed with fluorescently labeled partners. In a larger-scale test, a slide was printed containing 10,799 spots of protein G and one spot of specific bait (FKBP12–rapamycin-binding protein, FRP) and was probed with a mixture of BODIPY–FL–IgG and Cy5–FKBP. All of the protein G spots were bound by the BODIPY reagent whereas the single spot of FRP was detected by the FKBP. The challenge is to extend this system to permit the analysis of large sets of unique proteins.

A distinct type of protein array that has been utilized for differential protein expression analysis is the tissue microarray [20]. This type of array consists of cylinders of tissue 0.6 mm in diameter punched out from paraffin blocks. Grids of these cylinders are constructed by aligning cylinders from a number of samples (e.g. several thousand) into a recipient paraffin block, and serial sections are cut. Each section can then be analyzed by fluorescent *in situ* hybridization or immunohistochemistry. Although this technique obviously relies on choosing specific antigens for visualization, it does allow the concurrent examination of potentially thousands of specimens across a gradient of severity (e.g. benign, *in situ* carcinoma, invasive carcinoma and metastatic carcinoma). This technique has been used to examine potential tumor markers in bladder cancer, renal cell carcinoma and prostate cancer [21–23] and combined with cDNA microarray analysis to study glioblastomas [24].

Conclusions

The ProteinChip® proteomics platform provides an outstanding system for chip-based analysis of differential protein expression, given its low sample requirements and high-throughput format. Moreover, once a specific biomarker is found, the same platform may be used for diagnosis. This is particularly useful if a panel of biomarkers is more sensitive or specific than a single biomarker. Because this platform does not rely on protein conformation for detection, it has a great advantage over new chip platforms designed to bind recombinant proteins in an array-based manner analogous to DNA microarrays. In their current state, these systems suffer from a number of requirements: the identity of the protein placed at each position must be known, the recombinant protein must be soluble, the protein must be folded such that an appropriate domain for detection is displayed on the chip surface and a detection method must be either available or developed. To be useful for general protein profiling, future generations of these techniques will require adaptability in design and detection. It is likely that the next iteration of microarray-based protein chips will be best suited for the study of specific families of proteins, as family members share a common domain to which a detecting moiety can be synthesized. Enzymes make particularly good profiling targets, as the detection of their activity is a natural assay (as an example, see Update). Future generations of these techniques will likely rely on the expression of fusion proteins that incorporate one or more domains that can be used simultaneously to increase the solubility of the protein, orient the protein on the chip surface and act as a detection substrate (e.g. change in fluorescence wavelength when a ligand is bound). Molecular breeding will almost certainly play a role in the development of such domains, while advances in surface chemistry will be needed to increase the capacity of chip surfaces. Finally, improvements in the sensitivity and sequencing capabilities of mass spectrometry instrumentation will enable identification of the proteins that are upregulated or downregulated.

Update

Recently, a chip designed to assay kinase activity was described [25•]. Because the entire yeast genome has been

sequenced, these researchers could clone and express glutathione-S-transferase (GST) fusions of almost all the yeast kinases. The kinase-assay chip was constructed by first creating a mold over which a liquid silicone elastomer was poured. This was allowed to harden, was peeled away and was then placed on a glass slide. This array, which measured 28 mm × 14 mm, contained wells 1.4 mm in diameter and 300 μm deep. The potential substrates were covalently attached to the walls of the wells and incubated with recombinant kinase in the presence of radiolabeled ATP. After the kinase reaction was completed, the wells were washed and the chip exposed to either X-ray film or a phosphoimager. In this manner, each kinase could be tested against each of the different substrates. It is easy to envision modifications of this system to study other enzymatic activities such as GTP cleavage or protein lipidation, as well as to serve as a platform for studying the affects of drug candidates on such enzymatic activities.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Weinberger SR, Morris TS, Pawlak M: **Recent trends in protein biochip technology.** *Pharmacogenomics* 2000, in press.
2. Rich RL, Myszka DG: **Advances in surface plasmon resonance biosensor analysis.** *Curr Opin Biotechnol* 2000, **11**:54-61.
3. Figeys D: **Array and lab on a chip technology for protein characterization.** *Curr Opin Mol Ther* 1999, **1**:685-694.
4. Guetens G, Van Cauwenberghe K, De Boeck G, Maes R, Tjaden UR, van der Greef J, Highley M, van Oosterom AT, de Bruijn EA: **Nanotechnology in bio/clinical analysis.** *J Chromatogr B Biomed Sci Appl* 2000, **739**:139-150.
5. • Merchant M, Weinberger SR: **Recent advancements in surface enhanced laser desorption/ionization time-of-flight mass spectrometry.** *Electrophoresis* 2000, **21**:1164-1177.

Reviews the evolution of ProteinChip® technology and some of its other applications, including the study of protein–protein interactions.

6. Thulasiraman V, McCutchen-Maloney SL, Motin VL, Garcia E: **Detection and identification of virulence factors in *Yersinia pestis* using SELDI ProteinChip® system.** *Biotechniques* 2001, in press.
7. Chernushevich I, Ens W, Standing KG: **Orthogonal-injection TOFMS for analyzing biomolecules.** *Anal Chem* 1999, **71**:452A-461A.
8. Banks RE, Dunn MJ, Forbes MA, Stanley A, Pappin D, Naven T, Gough M, Harnden P, Selby PJ: **The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis – preliminary findings.** *Electrophoresis* 1999, **20**:689-700.
9. Paweletz CP, Gillespie JW, Ornstein DK, Simone NL, Brown MR, Cole KA, Wang Q-H, Huang J, Hu N, Yip T-T *et al.*: **Rapid protein display profiling of cancer progression directly from human tissue using a protein biochip.** *Drug Dev Res* 2000, **49**:34-42.
10. von Eggeling F, Davies H, Lomas L, Fiedler W, Junker K, Claussen U, Ernst G: **Tissue-specific microdissection coupled with ProteinChip® Array technologies: applications in cancer research.** *Biotechniques* 2000, **29**:1066-1070.
11. Wright GL Jr, Cazares LH, Leung S-M, Nasim S, Adam B-L, Yip T-T, Schellhammer PF, Gong L, Vlahou A: **ProteinChip surface enhanced laser desorption/ionization (SELDI) mass spectrometry: a novel protein biochip technology for detection of prostate cancer biomarkers in complex protein mixtures.** *Prostate Cancer and Prostatic Diseases* 2000, **2**:264-276.
12. Adam B-L, Davis JW, Cazares LH, Schellhammer PF, Wright GL Jr: **Identifying the signature proteins of prostate cancer in seminal plasma by SELDI affinity mass spectrometry.** *Proc Amer Assoc Cancer Res* 2000, **41**:564.
13. Vlahou A, Mendrinos S, Kondylis F, Schellhammer PF, Lynch DF, Wright GL Jr: **Identification of protein changes in bladder cancer patient urines by ProteinChip SELDI affinity mass spectrometry.** *Proc Amer Assoc Cancer Res* 2000, **41**:852-853.
14. Adam B-L, Davis JW, Xiao Z, Cazares LH, Schellhammer PF, Wright GL Jr: **Simultaneous measurement of free, complexed, and total PSA using a novel biochip multiplex SELDI mass spectrometry immunoassay.** *J Urol* 2000, **168**:180.
15. Xiao Z, Jiang X, Beckett ML, Wright GL Jr: **Generation of a baculovirus recombinant prostate-specific membrane antigen and its use in the development of a novel protein biochip quantitative immunoassay.** *Protein Expr Purif* 2000, **19**:12-21.
16. Emili AQ, Cagney G: **Large-scale functional analysis using peptide or protein arrays.** *Nat Biotechnol* 2000, **18**:393-397.
17. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al.*: **A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403**:623-627.
18. •• de Wildt RMT, Mundy CR, Gorick BD, Tomlinson IM: **Antibody arrays for high-throughput screening of antibody–antigen interactions.** *Nat Biotechnol* 2000, **18**:989-994.

First demonstration of a large-scale antibody array, with more than 18,000 antibodies placed on a 22 × 22 cm grid. This study also shows the reproducibility of such an array; reproducibility of any protein array is crucial but difficult to achieve.

19. • MacBeath G, Schreiber SL: **Printing proteins as microarrays for high-throughput function determination.** *Science* 2000, **289**:1760-1763.

Novel technique to print protein microarrays in a fashion analogous to DNA microarrays. The authors demonstrate the utility of such a system to identify binding partners for specific printed proteins.

20. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: **Tissue microarrays for high-throughput molecular profiling of tumor specimens.** *Nat Med* 1998, **4**:844-847.
21. Richter J, Wagner U, Kononen J, Fijan A, Bruderer J, Schmid U, Ackermann D, Maurer R, Alund G, Knonagel H, *et al.*: **High-throughput tissue microarray analysis of cyclin E gene amplification and overexpression in urinary bladder cancer.** *Am J Pathol* 2000, **157**:787-794.
22. Moch H, Schraml P, Bubendorf L, Mirlacher M, Kononen J, Gasser T, Mihatsch MJ, Kallioniemi OP, Sauter G: **High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma.** *Am J Pathol* 1999, **154**:981-986.
23. Bubendorf L, Kolmer M, Kononen J, Koivisto P, Mousses S, Chen Y, Mahlamaki E, Schraml P, Moch H, Willi N, *et al.*: **Hormone therapy failure in human prostate cancer: analysis by complementary DNA and tissue microarrays.** *J Natl Cancer Inst* 1999, **91**:1758-1764.
24. Sallinen SL, Sallinen PK, Haapsalo HK, Helin HJ, Helen PT, Schraml P, Kallioniemi OP, Kononen J: **Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques.** *Cancer Res* 2000, **60**:6617-6622.
25. • Zhu H, Klemic JF, Chang S, Bertone P, Casmayor A, Klemic KG, Smith D, Gerstein M, Reed MA, Snyder M: **Analysis of yeast protein kinases using protein chips.** *Nat Genet* 2000, **26**:283-289.

Novel chip-based strategy for studying a large number of kinases. This chip design has the potential to serve as a platform for assessing the specificity of a large number of enzymatic reactions or interaction combinations.

Towards a truly integrative biology through the functional genomics of yeast

Daniela Delneri*§, Francesco L Brancia*†# and Stephen G Oliver*‡

A complete library of mutant *Saccharomyces cerevisiae* strains, each deleted for a single representative of yeast's 6000 protein-encoding genes, has been constructed. This represents a major biological resource for the study of eukaryotic functional genomics. However, yeast is also being used as a test-bed for the development of functional genomic technologies at all levels of analysis, including the transcriptome, proteome and metabolome.

Addresses
*School of Biological Science, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK
†Michael Barber Centre for Mass Spectrometry, UMIST, Sackville Street, Manchester M60 1QD, UK
‡Corresponding author: SG Oliver; e-mail: steve.oliver@man.ac.uk
§e-mail: d.delneri@man.ac.uk
#e-mail: francesco.brancia@stud.umist.ac.uk

Current Opinion in Biotechnology 2001, **12**:87–91

0958-1669/01/$ – see front matter

Abbreviations

2DGE	two-dimensional gel electrophoresis
ICATs	isotope-coded affinity tags
MS	mass spectrometry
ORF	open reading frame

Introduction

The availability of complete genome sequences presents the opportunity of adopting a systems-based approach to biology that will allow the determination of how all the genes in a genome act and interact to produce a functioning organism. Such an approach demands that technologies are developed that allow analyses at the level of mRNAs (transcriptome), proteins (proteome) and low molecular weight intermediates (metabolome) to be carried out in as comprehensive a manner as possible. Moreover, high-throughput methods of generating defined mutants and assessing their phenotypes also are required. The full range of genomic technologies is currently being applied to a small number of model organisms. Of these, the yeast *Saccharomyces cerevisiae* is in the vanguard.

The three levels of functional genomic analysis are both qualitatively and quantitatively distinct. Messenger RNA molecules, the subject of transcriptome analysis, may be studied in a fully comprehensive manner using the massively parallel technique of hybridisation-array analysis. However, mRNA molecules are not functional entities within the cell, but simply transmitters of the instructions for synthesising proteins, and so transcriptome analyses only approach functionality in an indirect manner. While both proteins and metabolites represent true functional entities within cells, the analysis of the proteome is fraught with technical difficulties, and most techniques in current use are neither comprehensive in their scope, nor massively parallel in their execution. The main difficulty with metabolome analysis is conceptual, rather than technical. While it is technically feasible to simultaneously analyse several hundred metabolites at once, the relationship between the metabolome and the genome differs from that of the transcriptome or proteome in that it is indirect. Many genes may be involved in the biosynthesis or degradation of a single metabolite. Thus bioinformatic tools are required that permit prior knowledge of the metabolic impact of known genes to be exploited in elucidating the function of novel genes.

It is bioinformatics that holds the key to functional genomics, since it is clear that investigations at all the levels of 'omic' analysis, and also phenotypic studies, will be required in taking this integrative approach to biology. Thus, massive amounts of data, of qualitatively and quantitatively distinct types, must be integrated, compared and used to construct heuristic models of living systems. In this short review, we will attempt to look at all of the levels of genomic analysis and discuss the prospects for their use in gaining an integrative view of the workings of *S. cerevisiae*.

Genome

A major conceptual and practical problem for the systematic analysis of gene function in eukaryotes is that of genetic redundancy. Gene duplications could have occurred through a series of local events or by complete genome duplication [1,2]. The high level of redundancy, generated by a whole-genome duplication, is thought to have been reduced via deletions and chromosomal rearrangements, while sequence divergence and selection allowed the acquisition of new functions. Thus, much of the redundancy in the yeast genome may be more apparent than real, with identical, or almost identical, gene products fulfilling distinct physiological roles due to differential gene expression or the targeting of similar proteins to different cellular compartments. For example, Delneri *et al.* [3] showed that only one fully functional member of the aryl-alcohol dehydrogenase (*AAD*) gene family in *S. cerevisiae* responds to oxidative stress. Moreover, theoretical studies have indicated that the genetic robustness of this organism does not rely on gene redundancy [4•].

Considerable progress has been made in the analysis of yeast gene function using single open reading frame (ORF) deletion mutants. However, it must be appreciated that the failure of any given mutant to reveal a phenotype may be the consequence of either genetic redundancy or of effective homeostatic controls within the cell. Thus, more studies are

required in which entire gene families are deleted [5,6] or where phenotype is assessed quantitatively, rather than qualitatively [7,8]. The year 2000 saw the establishment of a collection of mutant yeast strains, each bearing a defined deletion in one of yeast's 6138 potential protein-encoding genes (for details, contact euroscarf@em.uni-frankfurt.de). Each deleted ORF is flanked by two 20 bp molecular barcodes that are unique for each deletion [9••]. These allow the parallel analysis of the phenotypes of a large number of deletants to be performed using competitive growth assays. Such analyses assume that the specific ORF replacement is the only genetic change in the deletant. However, the transformation is both mutagenic and recombinogenic, and competitive hybridisation-array studies have shown that deletants may be aneuploid for whole chromosomes or chromosomal segments [10].

Ross-Macdonald *et al.* [11•] have taken a different approach to the generation of a large collection of defined yeast mutants. They used a transposon to mutagenise a yeast clone library in *Escherichia coli*. Individual plasmids were then prepared and used to transform a diploid yeast strain, where each plasmid integrates at its corresponding chromosomal locus, replacing the endogenous copy of the gene. The structure of the transposon allowed them to insert a short haemagglutinin tag within the yeast ORF that may be used for immunolocalisation of the tagged proteins. This permitted more than 300 previously nonannotated ORFs to be identified and the localisation of their protein products to be determined. This collection of mutants has also been used to determine disruption phenotypes for about 8000 strains, using 20 different growth conditions.

Genome-wide expression analysis (see below) has been used to follow adaptive evolution in yeast. The global transcript profiles from mutants selected during aerobic growth in a glucose-limited chemostat were compared with that of their parental strain. Genes involved in glycolysis and the tricarboxylic acid cycle showed alterations in expression in all three independently evolved strains, indicating that increased fitness is acquired by reducing the percentage of glucose which is fermented and increasing that channelled to respiration [12].

Transcriptome

Hybridisation arrays are now used widely to study the effects of cell physiology, development biology, or genetic constitution on the global expression pattern of yeast. For instance, yeast genes regulated directly or indirectly by the transcriptional activator Yap1p have been identified. The Yap1p-binding site was not always found in the promoter region of the target genes, and such genes are presumably under the indirect control of the activator. As ever, it is important to carefully define and regulate cell physiology when carrying out these global analyses. Thus, the *RPI1* gene (which encodes a repressor of the Ras–cAMP pathway) was found to be downregulated by Yap1p during the exponential growth phase, but upregulated in the stationary phase or following oxidative stress [13]. Careful control of cell physiology was employed by Ter Linde *et al.* [14] in their investigation of adaptation to aerobiosis and anaerobiosis in *S. cerevisiae*. About 93% of the ORFs analysed were expressed during both aerobic and anaerobic conditions, but about 140 and 219 genes showed a threefold higher transcription level under anaerobic and aerobic conditions, respectively.

Hughes *et al.* [15•] have recently constructed a 'compendium' of expression profiles corresponding to 300 diverse mutation or chemical treatments. In this way, the authors were able to assign functions to uncharacterised ORFs determining the biochemical and cellular pathways affected, via pattern matching. The compendium also was used to determine a novel target of the drug dyclonine, a sodium-channel blocker.

Proteomics

Although the yeast proteome (as it came to be known) has been studied, using two-dimensional gel electrophoresis (2DGE) since the late 1970s, progress in identifying the proteins contained within the spots on such gels has been disappointingly slow [16]. The availability of the complete yeast genome sequence and the development of 'soft' ionisation techniques for mass spectrometry (MS) have done much to speed up spot assignments, but improvements in MS and bionformatic techniques are still required. Recent advances include a method for *de novo* peptide sequencing [17] that improves fragmentation efficiency in post-source decay experiments, the use of guanidination to improve the signal response of carboxy-terminal lysine peptides [18], and the combination of the latter with Edman-type derivatisation [19]. This double-derivatisation approach offers a dual advantage, as both the total number of peptide masses available for database searching is increased, and the search space is reduced due the identification of the carboxy- and amino-terminal amino acids [20].

Yates Jr and co-workers [21] have developed a method for protein separation that is an attractive alternative to 2DGE. Known by the acronym MudPIT (multidimensional protein identification technique), it preserves protein–protein interactions and so facilitates the analysis of multisubunit complexes. MudPIT combines reversed-phase liquid chromatography with either cation-exchange or size-exclusion chromatography prior to analysis with electrospray tandem MS. This method was used to dissect the yeast 80S ribosomal subunit and resulted in the identification of an additional 11 proteins that had not previously been detected using the 2DGE approach.

An alternative approach to the identification of protein–protein interactions was applied in the analysis of the yeast nuclear pore complex (Nup85p), [22]. The complex was isolated using an affinity tag, crosslinked, and the resulting proteins pairs were resolved in sodium dodecyl sulphate-polyacrylamide gel electrophoresis and identified via matrix-assisted laser desorption/ionisation (MALDI)–MS, providing a model of the spatial organisation of the complex.

Gygi *et al.* [23••] combined sequence identification of the components in complex mixtures and accurate relative quantification. It exploits a novel class of chemical reagents, known as isotope-coded affinity tags (ICATs), together with tandem MS. The ICAT reagents contain both a biotin moiety and a thiol group, which is covalently attached to each cysteinyl residue in every protein. These two moieties are joined by a linker that contains either normal hydrogen or its heavy isotope, deuterium. Two protein extracts, obtained from yeast cultures grown on either ethanol or galactose, were compared by treating them with the isotopically light and heavy ICAT reagents, respectively.

A comprehensive analysis of protein–protein interaction in yeast has been undertaken recently by Uetz *et al.* [24•]. They employed two complementary strategies to exploit the yeast two-hybrid system for high-throughput analyses. The array approach (using pairwise crosses) yields more positive interactions, while the high-throughput library approach is more comprehensive, but less productive. Of the 12 'baits' that gave positive interactions with both screens, 48 possible partners were identified by the array approach, against only 14 in the library screen.

The relationship between mRNA and protein expression levels was investigated recently in a genome-wide context. Two groups of scientists made independent analyses comparing, under a given set of physiological conditions, the amount of proteins from a two-dimensional protein gel with the corresponding amount of transcripts calculated from the published serial analysis of gene expression (SAGE) analysis. Futcher *et al.* [25] found a satisfactory correlation between mRNA abundance, protein abundance and codon bias (measured in glucose and ethanol media). On the other hand, Gygi *et al.* [26] found that, for some genes, the mRNA abundance and the corresponding protein levels varied by more than 20-fold. According to this second study, the yeast proteome could not yet be predicted from the simple deduction of the transcript level because of the limits of the current approaches for quantitative analysis of protein levels.

Metabolomics

There are two major approaches to the assignment of gene function via metabolic analyses. One, which provides a direct link to the genome, is to uncover the biochemical reactions catalysed by enzymes encoded by genes of unknown function. Such an approach has been adopted by Martzen and co-workers: they developed a genomic strategy to identify yeast genes specifying biochemical activities by constructing a library of plasmids expressing glutathione *S*-transferase tagged yeast proteins. Using this strategy, they were able to identify proteins with novel biochemical activities [27•]. The problem of such an approach, as with the assignment of function via sequence homology, is that it attributes mechanism, rather than biological function, and is completely context-independent.

An alternative approach is to study the change in the cell's metabolic profile (or metabolome) which results from the deletion or overexpression of a given gene, and to assign function by comparing the metabolome change that result from the deletion of unknown genes with those that occur due to similar manipulations of known genes. This approach has been applied to *S. cerevisiae* and has been termed FANCY (for functional analysis by co-responses in yeast) [28•]. It involves a comprehensive analysis of cellular metabolites using either MS or nuclear magnetic resonance spectroscopy, combined with sophisticated chemometric analysis of the resulting spectra. The approach is able to reveal a phenotype for gene deletions that have no measurable effect on cell growth and cluster together metabolome profiles resulting from the deletion of genes affecting similar domains of metabolism. Yeast has a rather a limited range of metabolites, but a similar approach has been successfully employed with the plant, *Arabidopsis thaliana* [29•], which produces a much more complex range of metabolites.

Bioinformatics

Thousands of data points are accumulated in a single genome-wide expression experiment. In this context, the role of bioinformatics becomes essential in order to make biological sense out of the data and to assign functions to uncharacterised coding regions. An obvious approach to assigning gene function from transcriptome data is to group together genes with similar expression profiles [30]. Such methods analyse patterns of gene activity in an 'unsupervised' fashion. That is, without recourse to a training set of data relating to genes of well-known function and regulatory pattern [31]. Recently, Brown *et al.* [32] introduced a method for the identification of functionally related yeast proteins based on the theory of support vector machines, which represent a supervised learning technique that exploits prior knowledge of gene function to identify clusters. An alternate approach that exploits genetic programming has been published recently and appeared to be even more successful in that it proved able to learn the class of helix-turn-helix proteins, which include the transcription factors [33].

A different approach to the cluster analysis has been developed by Marcotte *et al.* [34••]. They grouped proteins by 'experimental data', 'related metabolic function', 'related phylogenetic profiles', 'rosetta stone method' (which links individual proteins whose homologues, in other organisms, are combined into a single multifunctional complex), and 'correlated mRNA expression'. Using these methods, they found a total of 93,000 pairwise links between functionally related yeast proteins, allowing the assignment of a general function to more than half the uncharacterised ORFs. Such an approach is an excellent example of the benefits of integrating data from all levels of functional genomic analysis. These integrative approaches require database structures that have sufficient breadth and flexibility to allow complex queries to be made over the qualitatively and quantitatively different datasets represented by the 'omes'. An appropriate object data model has been constructed recently that permits the integration of genome, transcriptome and proteome data for yeast [35].

Conclusions

The fact that the genome sequences of even well-characterised organisms contained about 40% of genes whose function had neither been established nor predicted was a shock to many biologists. The challenge was to assign a function to each of these novel genes. However, the comprehensive methods of analysis that are used to pursue these assignments have revealed that our view of biological function is rather one-dimensional. The hope now is that the analysis of genome, transcriptome, proteome and metabolome, as well as the phenotype, will allow a much more integrative view of biology at the level of the whole organism. In the early stages of such a process, single-celled organisms offer the advantages of simplicity combined with genetic and physiological malleability. The experimental methods and theoretical framework established using an organism such as yeast should provide a firm foundation for an integrative biology of human beings and their domestic plants and animals.

Acknowledgements

Functional genomics research in the authors' laboratory is supported by the Biotechnology and Biological Sciences Research Council, the Wellcome Trust, and the European Commission. F Brancia thanks the BBSRC and Nycomed-Amersham for a CASE Studentship.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Wolfe KH, Shields DC: **Molecular evidence for an ancient duplication of the entire yeast genome.** *Nature* 1997, **387**:708-713.

2. Seoighe C, Wolfe KH: **Updated map of duplicated regions in the yeast genome.** *Gene* 1999, **238**:253-261.

3. Delneri D, Gardner DCJ, Oliver SG: **Analysis of the seven-member *AAD* gene set demonstrates that genetic redundancy in yeast may be more apparent than real.** *Genetics* 1999, **153**:1591-1600.

4. • Wagner A: **Robustness against mutations in genetic networks of yeast.** *Nat Genet* 2000, **24**:355-361.

Using functional genomics data from *S. cerevisiae*, the author demonstrates that robustness against mutation is mainly due to interactions of unrelated genes rather than gene redundancy in itself.

5. Delneri D, Tomlin GC, Wixon JL, Hutter A, Sefton M, Louis EJ, Oliver SG: **Exploring redundancy in the yeast genome: an improved strategy for use of the cre-loxP system.** *Gene* 2000, **252**:127-135.

6. Llorente B, Fairhead C, Dujon B: **Genetic redundancy and gene fusion in the genome of the baker's yeast *Saccharomyces cerevisiae*: functional characterization of a three-member gene family involved in the thiamine biosynthetic pathway.** *Mol Microbiol* 1999, **32**:1140-1152.

7. Giaever G, Shoemaker DD, Jones TW, Liang H, Winzeler EA, Astromoff A, Davis RW: **Genomic profiling of drug sensitivities via induced haploinsufficiency.** *Nat Genet* 1999, **21**:278-283.

8. Baganz F, Hayes A, Farquhar R, Butler PR, Gardner DCJ, Oliver SG: **Quantitative analysis of yeast gene function using competition experiments in continuous culture.** *Yeast* 1998, **14**:1417-1427.

9. •• Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H *et al.*: **Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis.** *Science* 1999, **285**:901-906.

Using a PCR-mediated gene disruption strategy, the authors constructed ~6900 *S. cerevisiae* strains with a precise deletion of more than one-third of the open reading frames in the genome. In addition, two molecular barcodes were introduced into each deletion strain to allow a large number of deletion strains to be analysed in parallel in competitive growth assays.

10. Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ *et al.*: **Widespread aneuploidy revealed by DNA microarray expression profiling.** *Nat Genet* 2000, **25**:333-337.

11. • Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, Jansen R, Cheung KH, Sheehan A, Symoniatis D, Umansky L *et al.*: **Large-scale analysis of the yeast genome by transposon tagging and gene disruption.** *Nature* 1999, **402**:413-418.

The authors used a transposon-based method for the genome-wide analysis of disruption phenotypes, gene expression and protein localisation. A yeast genomic DNA library was mutagenised in *E. coli* with a mini-transposon (mTn) containing the *URA3* marker and *lacZ* reporter gene flanked by *loxP* sites, and three copies of haemagglutinin. This collection of mutants has been used to determine disruption phenotypes for about 8,000 strains using 20 different growth conditions. Moreover, 300 previously nonannotated open reading frames were identified and analysed by indirect immunofluorescence.

12. Ferea TL, Botstein D, Brown PO, Rosenzweig RF: **Systematic changes in gene expression patterns following adaptive evolution in yeast.** *Proc Natl Acad Sci USA* 1999, **96**:9721-9726.

13. Dumond H, Danielou N, Pinto M, Bolotin-Fukuhara M: **A large-scale study of Yap1p-dependent genes in normal aerobic and H_2O_2-stress conditions: the role of Yap1p in cell proliferation control in yeast.** *Mol Microbiol* 2000, **36**:830-845.

14. ter Linde JJ, Liang H, Davis RW, Steensma HY, van Dijken JP, Pronk JT: **Genome-wide transcriptional analysis of aerobic and anaerobic chemostat cultures of *Saccharomyces cerevisiae*.** *J Bacteriol* 1999, **181**:7409-7413.

15 • Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD *et al.*: **Functional discovery via a compendium of expression profiles.** *Cell* 2000, **102**:109-126.

Using a 'compendium database' of expression profiles from 300 different mutations and chemical treatments in *S. cerevisiae*, the authors were able to identify the cellular pathways affected via pattern matching.

16. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R: **Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.** *Proc Natl Acad Sci USA* 2000, **97**:9390-9395.

17. Keough T, Youngquist RS, Lacey MP: **A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry.** *Proc Natl Acad Sci USA* 1999, **96**:7131-7136.

18. Brancia FL, Oliver SG, Gaskell SJ: **Improved matrix-assisted laser desorption/ionization mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides.** *Rapid Commun Mass Spectr* 2000, **14**:2070-2073.

19. Summerfield SG, Bolgar MS, Gaskell SJ: **Promotion and stabilization of b(1) ions in peptide phenylthiocarbamoyl derivatives: Analogies with condensed-phase chemistry.** *J Mass Spectrom* 1997, **32**:225-231.

20. Brancia FL, Butt A, Benyon RJ, Hubbard SJ, Gaskell SJ, Oliver SG: **A combination of chemical derivatisation and improved bioinformatic tools optimises protein identification for proteomics.** *Electrophoresis* 2001, in press.

21. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III: **Direct analysis of protein complexes using mass spectrometry.** *Nat Biotechnol* 1999, **17**:676-682.

22. Rappsilber J, Siniossoglou S, Hurt EC, Mann M: **A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry.** *Anal Chem* 2000, **72**:267-275.

23 •• Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nat Biotechnol* 1999, **17**:994-999.

This method is based on a class of new chemical reagents termed isotope-coded affinity tags and tandem mass spectrometry. The relative quantification exploits stable isotope dilution techniques and enabled comparison of protein expression in the yeast *S. cerevisiae*, using either ethanol or galactose as a carbon source.

24 • Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al.*: **Protein–protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403**:623-627.

A genomic scale analysis of protein–protein interactions in yeast was undertaken by the two-hybrid method, with more than 2500 interactions described.

25. Futcher B, Latter GI, Monardo P, McLaughlin CS, Garrels JI: **A sampling of the yeast proteome.** *Mol Cell Biol* 1999, **19**:7357-7368.

26. Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Mol Cell Biol* 1999, **19**:1720-1730.

27 • Martzen MR, McCraith SM, Spinelli SL, Torres FM, Fields S, Grayhack EJ, Phizicky EM: **A biochemical genomics approach for identifying genes by the activity of their products.** *Science* 1999, **286**:1153-1155.
The authors developed a genomic strategy to identify yeast genes specifying biochemical activities: a set of ~6144 strains carrying a plasmid expressing different open reading frames (ORFs) fused with a glutathione *S*-transferase (GST) tag were generated. The GST-ORFs were purified in 64 pools of 96 fusions each. Every pool was assayed for a particular activity, and to determine, within the positive pool, the source strain responsible for that activity. GST-ORFs from each of the eight rows and 12 columns of strains from the microtitre plate were tested. Using this strategy, they identified a cyclic phosphodiesterase, and Appr-1′-p-processing activity and a cytochrome c methyltransferase.

28 • Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ *et al.*: **Functional genomics via the metabolome: a FANCY approach to characterising mutations with a silent phenotype.** *Nat Biotechnol* 2001, **19**:45-50.
The authors studied the change in the yeast metabolome as result of gene deletion. Clustering together metabolome profiles resulting from the deletion of genes affecting similar domains of metabolism, the authors were able to reveal a phenotype for gene deletions that has no measurable effect on yeast cell growth, and to assign the gene to a functional domain.

29 • Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L: **Metabolite profiling for plant functional genomics.** *Nat Biotechnol* 2000, **18**:1157-1161.
Using gas chromatography/mass spectrometry, the authors were able to quantify more than 300 compounds from *Arabidopsis thaliana* and to assign a chemical structure to about half of them. They compared four genotypes, showing that each of them possess a distinct metabolic profile.

30. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.

31. Kell DB, King RD: **On the optimization of classes for the assignment of unidentified reading frames in functional genomics programmes: the need for machine learning.** *Trends Biotechnol* 2000, **18**:93-98.

32. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr., Haussler D: **Knowledge-based analysis of microarray gene expression data by using support vector machines.** *Proc Natl Acad Sci USA* 2000, **97**:262-267.

33. Gilbert RJ, Rowland JJ, Kell DB. **Genomic computing: explanatory modelling for functional genomics.** In *Proceedings of the Genetic and Evolutionary Computation Conference.* Edited by Whitely D, Goldberg D, Cantú-Paz E, Spector L, Permee I, Beyer HG. San Francisco: Morgan Kaufmann; 2000:551-557.

34 •• Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D: **A combined algorithm for genome-wide prediction of protein function.** *Nature* 1999, **402**:83-86.
The authors developed a new method to find protein–protein interactions. They linked the proteins by 'correlated evolution', 'correlated mRNA expression' and 'patterns of domain evolution', finding a total of 93,000 pairwise links between functionally related yeast proteins, and assigning a general function to more than half uncharacterised open reading frames.

35. Paton NW, Khan SA, Hayes A, Moussouni F, Brass A, Eilbeck K, Goble CA, Hubbard SJ, Oliver SG: **Conceptual modelling of genomic information.** *Bioinformatics* 2000, **16**:548-557.

Global Analysis of Protein Activities Using Proteome Chips

Heng Zhu,[1] Metin Bilgin,[1] Rhonda Bangham,[1] David Hall,[2] Antonio Casamayor,[1] Paul Bertone,[1] Ning Lan,[2] Ronald Jansen,[2] Scott Bidlingmaier,[2] Thomas Houfek,[3] Tom Mitchell,[3] Perry Miller,[4] Ralph A. Dean,[3] Mark Gerstein,[2] Michael Snyder[1,2]*

To facilitate studies of the yeast proteome, we cloned 5800 open reading frames and overexpressed and purified their corresponding proteins. The proteins were printed onto slides at high spatial density to form a yeast proteome microarray and screened for their ability to interact with proteins and phospholipids. We identified many new calmodulin- and phospholipid-interacting proteins; a common potential binding motif was identified for many of the calmodulin-binding proteins. Thus, microarrays of an entire eukaryotic proteome can be prepared and screened for diverse biochemical activities. The microarrays can also be used to screen protein-drug interactions and to detect posttranslational modifications.

A daunting task after a genome has been fully sequenced is to understand the functions, modification, and regulation of every encoded protein (*1*). Currently, much effort is devoted toward studying gene, and hence protein, function and regulation by analyzing mRNA expression profiles, gene disruption phenotypes, two-hybrid interactions, and protein subcellular localization (*2*). Although these studies are useful, much information about protein function can be derived from the analysis of biochemical activities (*3–7*). In principle, the biochemical activities of proteins can be systematically probed by producing proteins in a high-throughput fashion and analyzing the functions of hundreds or thousands of protein samples in parallel using protein microarrays (*5, 6, 8*). Major hurdles in screening an entire proteome array have been the ability to generate the necessary expression clones and also the expression and purification of proteins in a high-throughput fashion.

We have constructed a yeast proteome microarray containing approximately 80% yeast proteins and screened it for a number of biochemical activities. We first built a high-quality collection of 5800 yeast open reading frames (ORFs) (93.5% of the total) cloned into a yeast high-copy expression vector using recombination cloning (*9*). The yeast proteins are fused to glutathione *S*-transferase polyhistidine (GST-HisX6) at their NH_2-termini and expressed in yeast using the inducible *GAL1* promoter (*5, 9*). The yeast expression strains contain individual plasmids in which the correct yeast ORFs have been shown to be fused in-frame to GST by DNA sequencing. The proteins were expressed in yeast to help ensure that the proteins were modified and folded properly. Using a 96-well format, 1152 samples were purified at once from yeast extracts using glutathione-agarose beads (*10*). We included 0.1% Triton in the lysis buffer and washes to ensure that the purified proteins were free of lipids. The quality and quantity of the purified proteins were monitored using immunoblot analysis of 60 random samples (Fig. 1A). Greater than 80% of the strains produced detectable amounts of fusion proteins of the expected molecular weight.

To prepare the proteome chips, we printed 6566 protein preparations, representing 5800 different yeast proteins, in duplicate onto glass slides using a commercially available microarrayer. Our initial experiments used aldehyde-treated microscope slides (*6*) in

[1]Department of Molecular, Cellular, and Developmental Biology and [2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. [3]Fungal Genomics Laboratory, North Carolina State University, Campus Box 7251, Raleigh, NC 27695–7251, USA. [4]Department of Anesthesiology, Yale University, New Haven, CT 06520, USA.

*To whom correspondence should be addressed. E-mail: michael.snyder@yale.edu

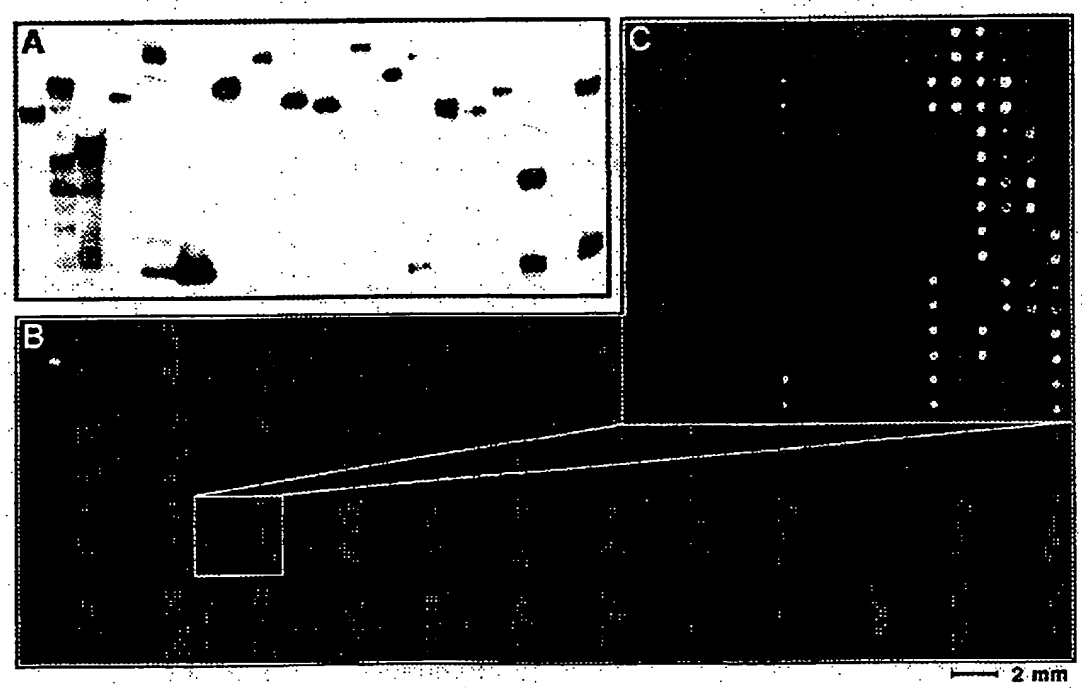


Fig. 1. GST::yeast proteins were purified in a 96-well format. (A) Sixty samples were examined by immunoblot analysis using anti-GST; 19 representative examples are shown. Greater than 80% of the preparations produce high yields of fusion protein. (B) 6566 protein samples representing 5800 unique proteins were spotted in duplicate on a single nickel-coated microscope slide. The slide was probed with anti-GST (*10*). (C) An enlarged image of one of the 48 blocks is depicted to the right of the proteome chip.

which fusion proteins attach to the surface through primary amines at their NH_2-termini or other residues of the protein. In subsequent experiments, we spotted proteins onto nickel-coated slides, in which the fusion proteins attach through their HisX6 tags and presumably uniformly orient away from the surface. Although both slides were successful, the nickel-coated slides gave superior signals for our particular protein preparations (Fig. 1B).

To determine how much fusion protein was covalently attached to different glass surfaces and the reproducibility of the protein attachment, we probed the chips with antibodies to GST (anti-GST). More than 93.5% of the protein samples gave signals significantly above background (i.e., greater than 10 fg of protein), and 90% of the spots contained 10 to 950 fg of protein. Our results also demonstrate that it is feasible to spot, with excellent resolution, 13,000 protein samples in one-half the area of a standard microscope slide (Fig. 1C). To test the reproducibility of the protein spotting, we compared the signals from each pair of duplicated spots with one another; 95% of the signals were within 5% of the average (*10*).

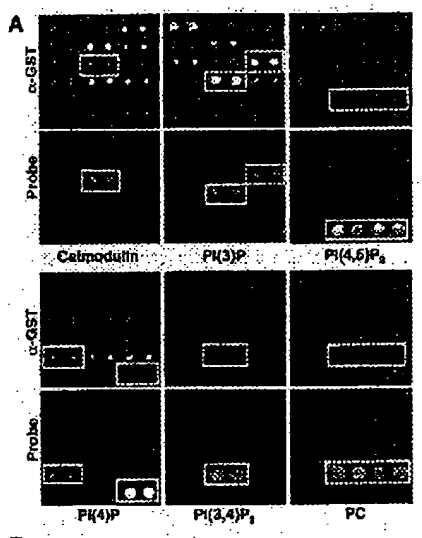


Fig. 2. (**A**) Examples of different assays on the proteome chips. Proteome chips containing 6566 yeast proteins were spotted in duplicate and incubated with the biotinylated probes indicated. The positive signals in duplicate (green) are in the bottom row of each panel; the top row of each panel shows the same yeast protein preparations of a control proteome chip probed with anti-GST (red). The upper panel shows the amounts of GST fusion proteins as detected by the anti-GST (red). (**B**) A putative calmodulin-binding motif (*32*) is shown, which was identified by searching for amino acid sequences that are shared by the different calmodulin targets (*10*). Fourteen of 39 positive proteins share a motif whose consensus is (I/L)QXK(K/X)GB, where X is any residue and B is a basic residue. The size of the letter indicates the relative frequency of the amino acid indicated.

The proteome chips were tested by probing for several protein-protein interactions and protein-lipid interactions. To test for protein-protein interactions, the yeast proteome was probed with biotinylated calmodulin in the presence of calcium (*11*). Calmodulin is a highly conserved calcium-binding protein involved in many calcium-regulated cellular processes and has many known partners (*12*). The bound biotinylated protein was detected using Cy3-labeled streptavidin. As a control, we also probed with Cy3-labeled streptavidin alone. These studies identified six known calmodulin targets (Fig. 2A): Cmk1p and Cmk2p are the type I and II calcium/calmodulin-dependent serine/threonine protein kinases (*12*), Cmp2p is one of the two yeast calcineurins (*13*), Dst1p plays a role in transcription elongation (*14*), Myo4p is a class V myosin heavy chain (*15*), and Arc35p is a component of the Arp2/3 actin-organizing complex (*16*). Arc35p was recently shown to interact with calmodulin in a two-hybrid study (*17*); thus, our data confirm that Arc35p and calmodulin interact in vitro. Of the six known calmodulin targets that we did not detect, two are not in our collection and the rest were not detectable in the GST probing experiments. In addition to known partners, the calmodulin probe identified 33 additional potential partners. These include many different types of proteins [supplementary table 1 (*10*)], consistent with a role for calmodulin in many diverse cellular processes.

Sequence searching (*5*) revealed that 14 of the 39 calmodulin-binding proteins contain a motif whose consensus is (I/L)QXXK(K/X)GB, where X is any residue and B is a basic residue (Fig. 2B). A related sequence in myosins, IQXXXXKXXXR, has been shown previously to bind calmodulin (*18*). Thus, we demonstrate that the domain is found in many calmodulin-binding proteins. Presumably the other targets that lack this motif have other calmodulin-binding sequences (*10*).

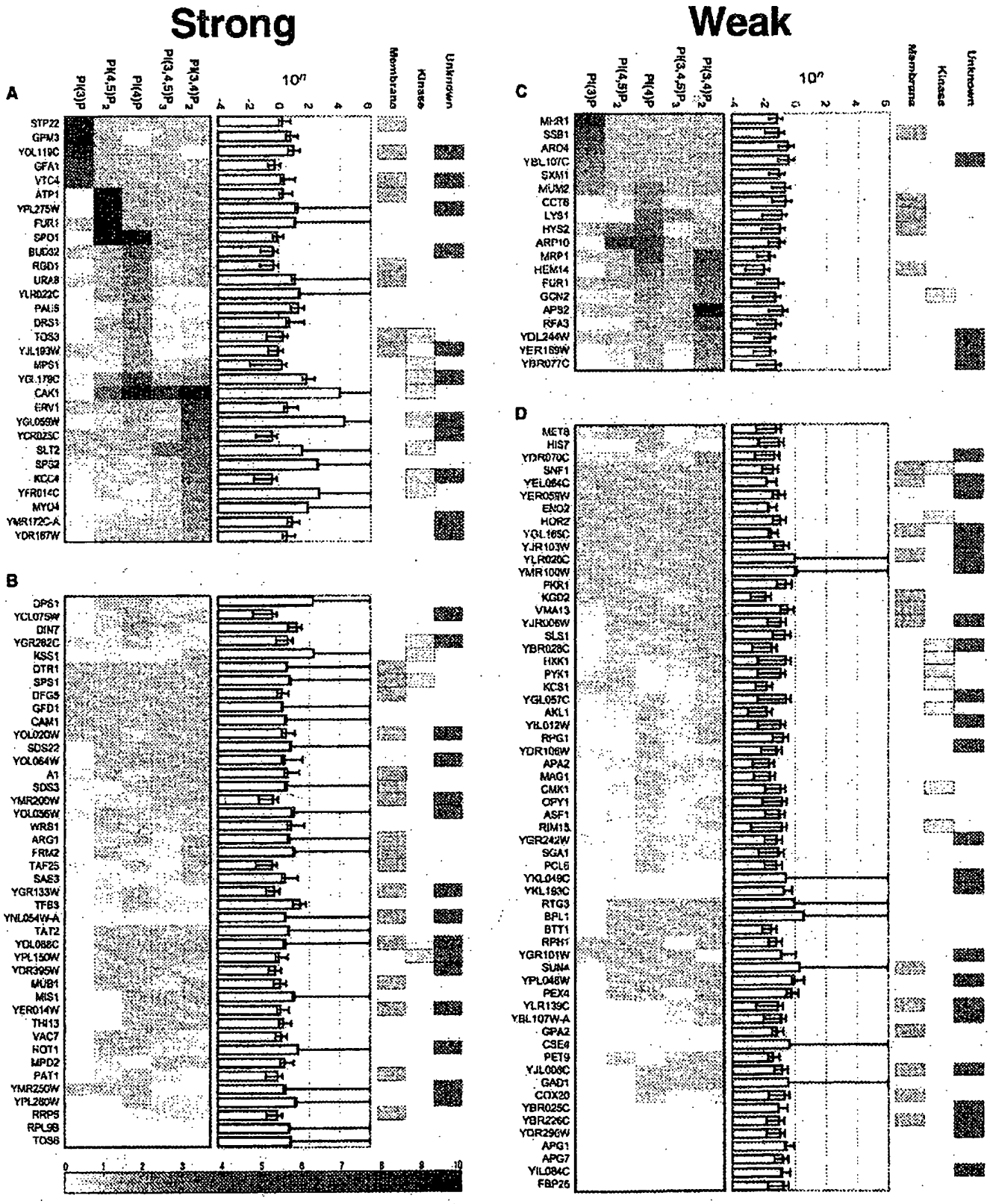
In addition to the calmodulin-binding targets, we also identified one protein, Pyc1p, that bound Cy3-labeled streptavidin. Pyc1p encodes a pyruvate carboxylase 1 homolog that contains a highly conserved biotin attachment region (*19*). Thus, as predicted by its sequence, Pyc1p is biotinylated in vivo. With appropriate detection assays, we expect that proteome chips can identify many types of posttranslational modification of proteins.

To test whether proteome chips could be used to identify activities that might not be accessible by other approaches, such as protein-drug interactions and protein-lipid interactions, we screened for phosphoinositide (PI)–binding proteins. PIs are important constituents of cellular membrane and also serve as second-messengers that regulate diverse cellular processes, including growth, differentiation, cytoskeletal rearrangements, and membrane trafficking (*20*). Because they are often present only transiently and in low abundance within cells, PIs have not been characterized extensively, and little is known about which proteins bind different phospholipids (*20*).

Five types of PI liposomes and one liposome lacking PIs were used to probe the proteome chips. Each contains phosphatidylcholine (PC) with 1% (w/w) *N*-(biotinoyl)-1,2-dihexadecanoyl-*sn*-glycero-3-phosphoethanolamine, triethylammonium salt (biotin DHPE); the biotinylated lipid serves as a label that can be detected by Cy3-streptavidin (*21*). In addition to PC, the five other liposomes contain either 5% (w/w) PI(3)P, PI(4)P, PI(3,4)P_2, PI(4,5)P_2, or PI(3,4,5)P_3 (Fig. 2A). All of these phospholipids have been found in yeast except PI(3,4,5)P_3 (*20*).

The six liposomes identified a total of 150 different protein targets that produced signals significantly higher than the background; an algorithm was devised to assist in the identification of positive signals (*22*). Fifty-two (35%) of the lipid-binding proteins correspond to uncharacterized proteins. Of the 98 known proteins, 45 proteins are membrane-associated and either have, or are predicted to have, membrane-spanning regions (*23, 24*).

Fig. 3. Analysis of the PI lipid-binding proteins. To determine the PI-binding specificity of 150 positive proteins, their binding signals were normalized against the corresponding binding signals of PC. On the basis of the ratios (PI/PC), the proteins were grouped into four categories: (**A**) 30 strong and specific, (**B**) 43 strong and nonspecific, (**C**) 19 weak and specific, and (**D**) 58 weak and nonspecific PI-binding proteins. The green color intensity represents the PI/PC signal ratio as shown by the scale in the figure. The first column to the right of the PI/PC binding ratios indicates the maximum binding signal intensity (open boxes) and its confidence interval (solid horizontal lines); the numbers indicate the log of the values. Blue, yellow, light-yellow, and red boxes in columns to the right of confidence interval column identify membrane-associated proteins, protein kinases, other kinases, and uncharacterized ORFs, respectively.



This includes integral membrane proteins, those with lipid modifications [e.g., the glycosylphosphatidylinositol (GPI) anchor proteins Tos6p and Sps2p (*23*) and prenylated proteins (Gpa2p and the mating pheromone a-factor) (*25*)], as well as peripherally associated proteins [e.g., Kcc4p and Myo4p (*15, 26*)]. Eight others are involved in lipid metabolism (e.g., Bpl1p) or inositol ring phosphorylation (e.g., Kcs1p) or are predicted to be involved in membrane or lipid function (e.g., Ylr020cp has homology to triacylglycerol lipase). Of the 52 uncharacterized proteins, 13 (25%) are predicted to be associated with membranes (*24*) and others contain basic stretches, as might be expected for electrostatic interactions with negatively charged lipids. Surprisingly, 19 of the lipid-binding proteins are kinases, and 17 of these are protein kinases.

The phospholipid-binding proteins were sorted into whether they bound lipids strongly or weakly, on the basis of the phospholipid-binding signal relative to the amount of GST (Fig. 3) (*22*). We found that more (72%) of the strong lipid-binding proteins (Fig. 3, A and B) were characterized relative to the weakly binding proteins (54%) (Fig. 3, C and D) and more strong lipid-binding proteins are known or predicted to be membrane-associated, relative to the weaker binding proteins (Fig. 3, "Membrane" column). Interestingly, 13 of 17 of the protein kinases bind very strongly to the PIs. We further grouped the proteins by whether they preferentially bound one or more PIs over PC. One-hundred and one proteins bound to PC as well as or nearly as well as to the PIs (PI/PC < 1.3) (Fig. 3, B and D). However, 49 proteins bound to one or more PIs preferentially (PI/PC > 1.3) (Fig. 3, A and C). Analysis of the strong PI-binding proteins revealed that many of them specifically bound particular PIs. For example, Stp22p, which is required for vacuolar targeting of plasma membrane proteins such as Ste2p and Can1p, preferentially binds PI(3)P (*27*). Nine protein kinases specifically bind PI(4)P and PI(3,4)P_2 strongly and one binds these lipids weakly. Atp1p, a subunit of the F1-ATP synthase of the mitochondrial inner membrane, preferentially binds PI(4,5)P_2 (*28*). Sps2p, which is localized to the prospore membrane (*29*), also interacts specifically with PI(3,4)P_2. Preferential binding of Myo4p to PI(3,4)P_2 may be important for its interaction at the cell cortex and/or its regulation. No strong lipid-binding targets were found that specifically bound PI(3,4,5)P_3, although some proteins bound both this lipid and others (Fig. 3). These results demonstrate that many membrane-associated proteins, including integral membrane proteins and peripherally associated proteins, preferentially bind specific phospholipids in vivo.

Several proteins involved in glucose metabolism were identified as phospholipid-binding proteins. This includes (i) three enzymes involved in sequential glycolytic steps [phosphoglycerate mutase (Gpm3p), enolase (Eno2p), and pyruvate kinase (Cdc19p/Pyk1p)], (ii) hexokinase (Hxk1p), and (iii) two protein kinases (Snf1p and Rim15p). Although unexpected, previous studies indicate that some of these might interact with lipids. Hxk1p binds zwitterion micelles, which stimulate its activity (*30*), and Eno2p is secreted, suggesting an interaction with membranes (*31*). We speculate that either phospholipids regulate steps involved in glucose metabolism or many steps of glucose metabolism occur on membrane surfaces. In the latter case, the phospholipids would serve as a scaffold to efficiently carry out glycolytic steps.

Six proteins not expected to be involved in membrane function or lipid signaling, Rim15p, Eno2p, Hxk1p, Sps1p, Ygl059wp, and Gcn2p, were further tested for PI-binding using two types of standard assays (*30*). For three proteins, Rim15p, Eno2p, and Hxk1p, PI(4,5)P_2 liposomes were first adhered to a nitrocellulose membrane; different amounts of the GST fusion proteins and a GST control were used to probe the membrane, and bound proteins were detected using anti-GST. As shown in Fig. 4A, each yeast fusion protein tightly bound PI(4,5)P_2, whereas GST alone did not. We also carried out the reverse assay for GST fusion proteins of Rim15p, Sps1p, Ygl059wp, and Gcn2p (*30*). Different amounts of these purified proteins were spotted onto nitrocellulose and probed with the six different liposomes (Fig. 4B); the bound liposomes were detected using a horseradish peroxidase (HRP)–conjugated streptavidin. As with the microarrays, liposomes bound to each protein, but not the bovine serum albumin control. Sps1p bound all five PI-containing liposomes nearly equally. Rim15p, Gcn2p, and Ygl059wp exhibited different affinities to different liposomes (see Fig. 4B for Rim15p); PI(3)P, PI(4)P, and PI(3,4)P_2 bound strongest. In each case, a linear correlation between the binding signal and the level of protein was revealed (*10*). In summary, these results demonstrate that PI-binding proteins identified in the proteome array also bind lipids in conventional assays.

One concern about our experiments is that because proteins are purified from yeast, we might detect indirect interactions through associated proteins. Most of the interactions that we detect are expected to be directly or at least tightly associated with the protein of interest, because proteins were prepared using stringent conditions, and for the seven samples examined, contaminating bands were not detected using Coomassie staining. Another limitation is that properly folded extracellular domains and secreted proteins are likely to be underrepresented in our collection, because GST and a HisX6 tag are fused at the NH_2-terminus. Thus, proteins with a signal peptide may not be delivered to the secretory pathway and may not be folded or modified properly, although we did detect three signal peptide–containing proteins, suggesting that at least some are produced and contain functional domains. Another limitation is that not all interactions are detected because not all proteins are readily overproduced and purified in this high-throughput approach; we expect that 80% of the arrayed yeast proteins are full length and at reasonable levels for screening.

Regardless, the use of proteome chips has significant advantages over existing approaches. Random expression libraries are incomplete, the clones are often not full length, and the libraries are tedious to screen. A recent alternative approach is to generate defined arrays and screen them using a pooling strategy (*4*). The pooling strategy requires two steps, the actual number of proteins screened is not known, and the method does not work well when large numbers of reacting proteins exist, because each pool will test positive. Another method for detecting interactions is the two-hybrid approach (*2*), in which interactions are typically detected in the nucleus, thus limiting the types of interactions that can be detected. The advantage of the proteome chip approach is that a comprehensive set of individual proteins can be directly screened in vitro for a wide variety of activities, including protein-drug interactions, protein-lipid interactions, and enzy-

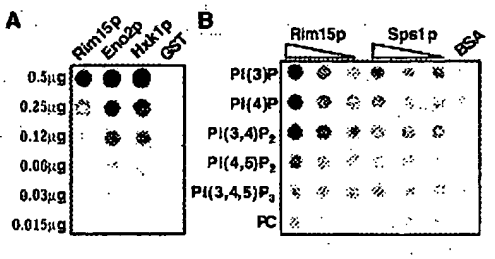


Fig. 4. Conventional methods confirm protein-lipid interactions detected by the proteome microarrays (*30*). (**A**) PI(4,5)P_2 liposomes were first adhered to a nitrocellulose membrane; a dilution series of Rim15p, Eno2p, and Hxk1p, and a GST control were used to probe the membrane. The bound proteins were detected using anti-GST and an ECL kit. (**B**) A reverse assay was carried out to test potential protein-lipid interactions. The proteins were prepared and spotted onto nitrocellulose filters in a dilution series and probed with the six different liposomes. As a control, the six liposomes were also added to the membrane. After extensive washing, the bound liposomes were detected using an HRP-conjugated streptavidin and an ECL kit.

matic assays using a wide range of in vitro conditions. Furthermore, once the proteins are prepared, proteome screening is significantly faster and cheaper. Using similar procedures, it is clearly possible to prepare protein arrays of 10 to 100,000 proteins for global proteome analysis in humans and other eukaryotes.

References and Notes

1. S. Fields, Y. Kohara, D. J. Lockhart, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8825 (1999); A. Goffeau *et al.*, *Science* **274**, 546 (1996).
2. P. Ross-Macdonald *et al.*, *Nature* **402**, 413 (1999); J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997); E. A. Winzeler *et al.*, *Science* **285**, 901 (1999); P. Uetz *et al.*, *Nature* **403**, 623 (2000); T. Ito *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1143 (2000).
3. H. Zhu, M. Snyder, *Curr. Opin. Chem. Biol.* **5**, 40 (2001).
4. M. R. Martzen *et al.*, *Science* **286**, 1153 (1999).
5. H. Zhu *et al.*, *Nature Genet.* **26**, 283 (2000).
6. G. MacBeath, S. L. Schreiber, *Science* **289**, 1760 (2000).
7. A. Caveman, *J. Cell Sci.* **113**, 3543 (2000).
8. P. Arenkov *et al.*, *Anal. Biochem.* **278**, 123 (2000).
9. D. A. Mitchell, T. K. Marshall, R. J. Deschenes, *Yeast* **9**, 715 (1993). The expression vector pEGH was created by inserting an RGS-HisX6 epitope tag between the GST gene and the polycloning site of pEG(KG). The yeast ORFs were cloned using the strategy described previously (5), except every step was done in a 96-well format. Plasmid DNAs confirmed by DNA sequencing were reintroduced into both yeast (Y258) and *E. coli* (DH5α). The library contains 5800 unique ORFs.
10. For details of 96-well format protein purification protocol, a full list of results from all the experiments, and the design of the positive identification algorithms, please visit our public Web site (http://bioinfo.mbb.yale.edu/proteinchip) and supplementary material at *Science* Online (www.sciencemag.org/cgi/content/full/1062191/DC1).
11. Biotinylated calmodulin (CalBiochem, USA) was added to the proteome chip at 0.02 μg/μl in phosphate-buffered saline (PBS) with 0.1 mM calcium and incubated in a humidity chamber for 1 hour at room temperature. Calcium (0.1 mM) was present in buffers in all subsequent steps. The chip was washed three times with PBS at room temperature (RT, 25°C). Cy3-conjugated streptavidin (Jackson IR, USA) (1:5000 dilution) was added to the chip and incubated for 30 min at RT. After extensive washing, the chip was spun dry and scanned using a microarray scanner; the data was subsequently acquired with the GenePix array densitometry software (Axon, USA).
12. S. S. Hook, A. R. Means, *Annu. Rev. Pharmacol. Toxicol.* **41**, 471 (2001).
13. M. S. Cyert, R. Kunisawa, D. Kaim, J. Thorner, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7376 (1991).
14. D. A. Stirling, K. A. Welch, M. J. Stark, *EMBO J.* **13**, 4329 (1994).
15. F. Bohl, C. Kruse, A. Frank, D. Ferring, R. P. Jansen, *EMBO J.* **19**, 5514 (2000); E. Bertrand *et al.*, *Mol. Cell* **2**, 437 (1998).
16. D. C. Winter, E. Y. Choe, R. Li, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7288 (1999).
17. C. Schaerer-Brodbeck, H. Riezman, *Mol. Biol. Cell* **11**, 1113 (2000).
18. K. Homma, J. Saito, R. Ikebe, M. Ikebe, *J. Biol. Chem.* **275**, 34766 (2000).
19. J. Menendez, J. Delgado, C. Gancedo, *Yeast* **14**, 647 (1998).
20. G. Odorizzi, M. Babst, S. D. Emr, *Trends Biochem. Sci.* **25**, 229 (2000); D. A. Fruman *et al.*, *Annu. Rev. Biochem.* **67**, 481 (1998); T. F. Martin, *Annu. Rev. Cell Dev. Biol.* **14**, 231 (2000); S. Wera, J. C. T. Bergsma, *FEMS Yeast Res.* **1**, 1406 (2001).
21. Liposomes were prepared using standard methods (30). Briefly, appropriate amounts of each lipid in chloroform were mixed and dried under nitrogen. The lipid mixture was resuspended in TBS buffer by vortexing. The liposomes were created by sonication. To probe the proteome chips, 60 μl of the different liposomes were added onto different chips. The chips were incubated in a humidity chamber for 1 hour at RT. After washing with TBS buffer for three times, Cy3-conjugated streptavidin (1:5000 dilution) was added to the chip and incubated for 30 min at RT.
22. Positives were identified using a combination of the GenePix software which computes a local intensity background for each spot and a series of algorithms we developed. Details can be found at http://bioinfo.mbb.yale.edu/proteinchip and at www.sciencemag.org/cgi/content/full/1062191/DC1.
23. M. C. Costanzo *et al.*, *Nucleic Acids Res.* **29**, 75 (2001).
24. M. Gerstein, *Proteins* **33**, 518 (1998).
25. K. Ansari *et al.*, *J. Biol. Chem.* **274**, 30052 (1999).
26. Y. Barral, M. Parra, S. Bidlingmaier, M. Snyder, *Genes Dev.* **13**, 176 (1999).
27. Y. Li, T. Kane, C. Tipper, P. Spatrick, D. D. Jenness, *Mol. Cell. Biol.* **19**, 3588 (1999).
28. I. Arnold *et al.*, *J. Biol. Chem.* **274**, 36 (1999).
29. S. Chu *et al.*, *Science* **282**, 699 (1998).
30. A. Casamayor *et al.*, *Curr. Biol.* **9**, 186 (1999); R. Guerra, M. L. Bianconi, *Biosci. Rep.* **20**, 41 (2000).
31. M. Pardo *et al.*, *Yeast* **15**, 459 (1999).
32. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. X indicates any residue.

33. We thank K. Nelson and S. Dellaporta for providing invaluable help. We also thank A. Kumar, G. Michaud, and C. Costigan for providing comments on the manuscript. This research is supported by grants from NIH. H.Z., A.C., and R.J. were supported by postdoctoral fellowships from the Damon Runyon–Walter Winchell Foundation, the Spanish Ministerio de Ciencia y Tecnologia, and by an IBM Graduate Research Fellowship, respectively.

2 May 2001; accepted 13 July 2001
Published online 26 July 2001;
10.1126/science.1062191
Include this information when citing this paper.

An fMRI Investigation of Emotional Engagement in Moral Judgment

Joshua D. Greene,[1,2*] R. Brian Sommerville,[1] Leigh E. Nystrom,[1,3] John M. Darley,[3] Jonathan D. Cohen[1,3,4]

The long-standing rationalist tradition in moral psychology emphasizes the role of reason in moral judgment. A more recent trend places increased emphasis on emotion. Although both reason and emotion are likely to play important roles in moral judgment, relatively little is known about their neural correlates, the nature of their interaction, and the factors that modulate their respective behavioral influences in the context of moral judgment. In two functional magnetic resonance imaging (fMRI) studies using moral dilemmas as probes, we apply the methods of cognitive neuroscience to the study of moral judgment. We argue that moral dilemmas vary systematically in the extent to which they engage emotional processing and that these variations in emotional engagement influence moral judgment. These results may shed light on some puzzling patterns in moral judgment observed by contemporary philosophers.

The present study was inspired by a family of ethical dilemmas familiar to contemporary moral philosophers (*1*). One such dilemma is the trolley dilemma: A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person instead of five. Ought you to turn the trolley in order to save five people at the expense of one? Most people say yes. Now consider a similar problem, the footbridge dilemma. As before, a trolley threatens to kill five people. You are standing next to a large stranger on a footbridge that spans the tracks, in between the oncoming trolley and the five people. In this scenario, the only way to save the five people is to push this stranger off the bridge, onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Ought you to save the five others by pushing this stranger to his death? Most people say no.

Taken together, these two dilemmas create a puzzle for moral philosophers: What makes it morally acceptable to sacrifice one life to save five in the trolley dilemma but not in the footbridge dilemma? Many answers have been proposed. For example, one might suggest, in a Kantian vein, that the difference between these two cases lies in the fact that in the footbridge dilemma one literally uses a fellow human being as a means to some independent end, whereas in the trolley dilemma the unfortunate person just happens to

[1]Center for the Study of Brain, Mind, and Behavior, [2]Department of Philosophy, 1879 Hall, [3]Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544, USA. [4]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15260, USA.

*To whom correspondence should be addressed. E-mail: jdgreene@princeton.edu

Cell, Vol. 92, 573–585, February 20, 1998, Copyright ©1998 by Cell Press

Orexins and Orexin Receptors: A Family of Hypothalamic Neuropeptides and G Protein-Coupled Receptors that Regulate Feeding Behavior

Takeshi Sakurai,[1,8] Akira Amemiya,[1,9] Makoto Ishii,[1] Ichiyo Matsuzaki,[1,10] Richard M. Chemelli,[1,2] Hirokazu Tanaka,[1] S. Clay Williams,[1] James A. Richardson,[3] Gerald P. Kozlowski,[4] Shelagh Wilson,[5] Jonathan R. S. Arch,[5] Robin E. Buckingham,[5] Andrea C. Haynes,[5] Steven A. Carr,[6] Roland S. Annan,[6] Dean E. McNulty,[6] Wu-Schyong Liu,[5] Jonathan A. Terrett,[5] Nabil A. Elshourbagy,[6] Derk J. Bergsma,[6] and Masashi Yanagisawa[1,7]

[1]Howard Hughes Medical Institute
Department of Molecular Genetics
[2]Department of Pediatrics
[3]Department of Pathology
[4]Department of Physiology
University of Texas Southwestern
Medical Center at Dallas
Dallas, Texas 75235–9050
[5]SmithKline Beecham Pharmaceuticals
Harlow, Essex CM19 5AD
United Kingdom
[6]SmithKline Beecham Pharmaceuticals
King of Prussia, Pennsylvania 19406

Summary

The hypothalamus plays a central role in the integrated control of feeding and energy homeostasis. We have identified two novel neuropeptides, both derived from the same precursor by proteolytic processing, that bind and activate two closely related (previously) orphan G protein-coupled receptors. These peptides, termed orexin-A and -B, have no significant structural similarities to known families of regulatory peptides. *prepro-orexin* mRNA and immunoreactive orexin-A are localized in neurons within and around the lateral and posterior hypothalamus in the adult rat brain. When administered centrally to rats, these peptides stimulate food consumption. *prepro-orexin* mRNA level is up-regulated upon fasting, suggesting a physiological role for the peptides as mediators in the central feedback mechanism that regulates feeding behavior.

Introduction

Members of the family of seven-transmembrane, G protein-coupled cell surface receptors (GPCRs) respond to a wide variety of signals, including photons, amines, lipids, peptides, and proteases. Ligand binding initiates intracellular signal transduction through the activation of heterotrimeric G proteins. All of the known small regulatory peptides (small peptide hormones and neuropeptides) exert their biological actions by acting on GPCRs. Recent efforts in genomics research have identified a large number of cDNA sequences that encode "orphan" GPCRs, i.e., putative GPCRs without known cognate ligands. Many of these orphan GPCRs are likely to be receptors for heretofore unidentified signaling molecules, including new peptide hormones and neuropeptides. GPCRs already represent the largest class of target molecules for drugs available in the clinic (Hardman et al., 1996). The orphan GPCRs therefore represent a fruitful resource for drug discovery (Stadel et al., 1997).

To approach these possibilities, we undertook a systematic biochemical search for endogenous peptide ligands for multiple orphan GPCRs, using a cell-based reporter system. These screening experiments led to the identification of a novel family of neuropeptides that bind to two closely related orphan GPCRs. We call these peptide ligands "orexins," after the Greek word orexis, which means appetite. The mRNA for the precursor of these peptides is abundantly and specifically expressed in the lateral hypothalamus and adjacent areas, a region classically implicated in the central regulation of feeding behavior and energy homeostasis (Oomura, 1980; Bernardis and Bellinger, 1993, 1996). These peptides stimulate food consumption when administered centrally, and their production is influenced by nutritional state of the animal. Here we report the identification and initial biological characterization of two orexins as well as their two receptors.

Results and Discussion

Purification and Structure of Orexins

Our general strategy was to screen high-resolution HPLC fractions of various tissue extracts for GPCR-agonist activity, using multiple orphan GPCR–expressing cell lines tested in parallel. For this purpose, we generated over 50 stable transfectant cell lines, each expressing a distinct orphan GPCR cDNA, which can be used as reporter cell lines in order to identify endogenous ligands for these receptors. We challenged the cells with HPLC fractions derived from tissue extracts and we monitored a number of signal transduction read-outs for heterotrimeric G protein activation. A potential problem in this approach is the presence of numerous endogenous GPCRs in the host cell line. To circumvent this problem in part, we assayed active fractions against at least 2–3 independent transfectant cell lines each expressing a different orphan GPCR. This enabled us to search for active fractions that activate only cells that express a specific receptor.

In this process, we discovered several reverse-phase HPLC fractions of rat brain extracts (Figure 1) that elicit a robust increase of cytoplasmic Ca^{2+} levels ($[Ca^{2+}]_i$) in HEK293 cells expressing an orphan GPCR originally

[7]To whom correspondence should be addressed.
[8]Present address: Institute of Basic Medical Sciences, University of Tsukuba, Ibaraki 3050006, Japan.
[9]Present address: First Department of Surgery, Osaka University Medical School, Osaka 5650871, Japan.
[10]Present address: Institute of Community Medicine, University of Tsukuba, Ibaraki 3050006, Japan.

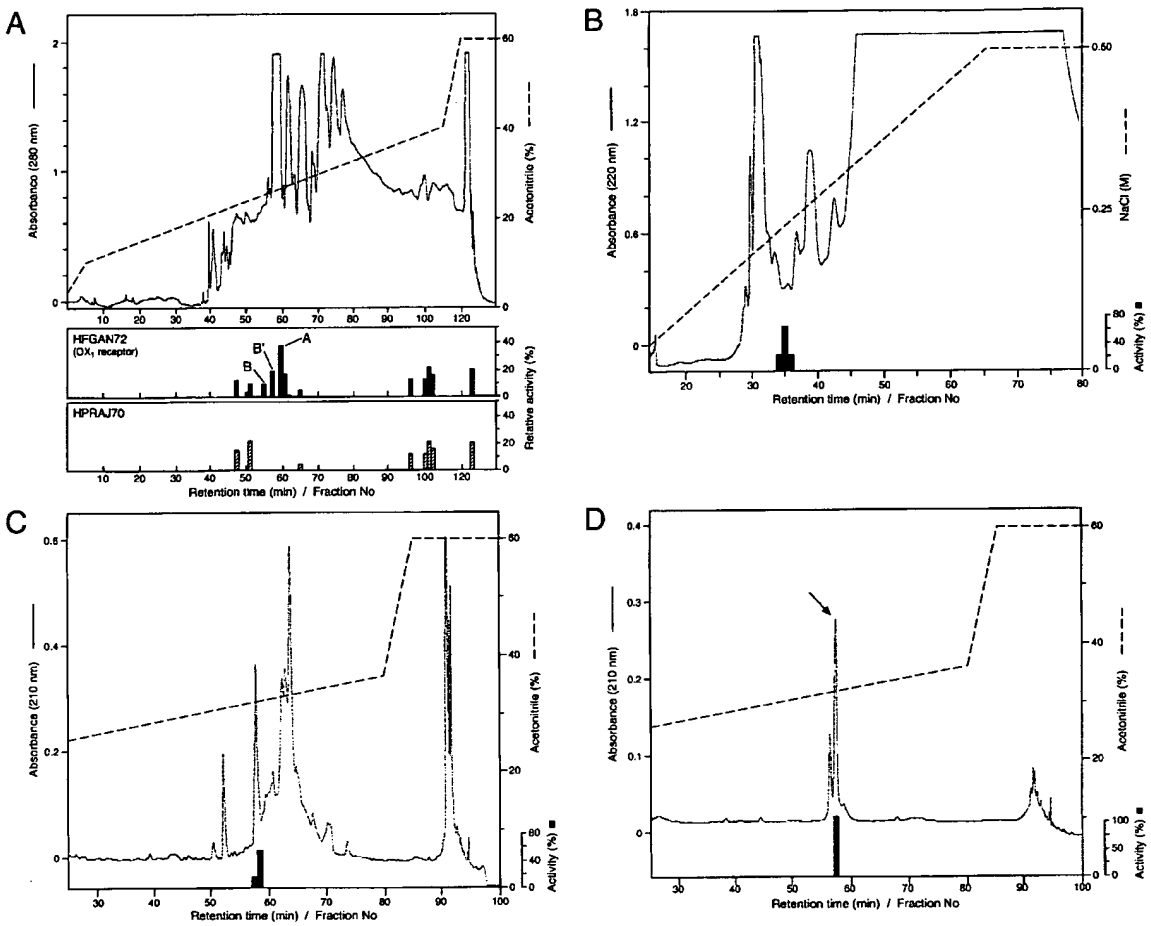


Figure 1. Purification of Endogenous Ligands for the Orphan Receptor HFGAN72 (OX_1R)

(A) Activation of the HFGAN72 receptor in transfected HEK293 cells by reverse-phase HPLC fractions of crude peptide extracts from rat brain. Top panel, chromatogram. Acetonitrile gradient and absorbance of eluted materials are indicated by solid and dotted lines, respectively. Middle and bottom panels, $[Ca^{2+}]_i$ transients in two HEK293 cell lines stably expressing unrelated orphan receptors *HFGAN72* and *HPRAJ70*, respectively. Peak increments in $[Ca^{2+}]_i$ evoked by designated HPLC fractions are normalized against the response induced by 100 nM endothelin-1 in the same reporter cell preparations and plotted as relative activity. Note that a number of fractions induced similar responses in both cell lines, presumably by acting on endogenous receptors in the host HEK293 cells. However, fractions 60–61 (marked as A; containing orexin-A) induced robust $[Ca^{2+}]_i$ transients in *HFGAN72*-transfected cells but not in *HPRAJ70*-transfected cells. Fractions 56 and 58 also exhibited weaker but distinct activities specific to HFGAN72 (marked as B and B'; containing orexin-B and orexin-B[3–28], respectively).

(B) Cation exchange HPLC of fractions 60–61 from Figure 1A. Bars show normalized magnitude of $[Ca^{2+}]_i$ in *HFGAN72*-expressing HEK293 cells as in Figure 1A.

(C) Reverse-phase HPLC of active fractions from Figure 1B.

(D) Final purification by reverse-phase HPLC at 40°C of active fractions in Figure 1C. Arrow, final orexin-A peak.

termed HFGAN72. This receptor was initially identified as an expressed sequence tag (EST) from human brain (see Figure 2C). A major activity peak (marked "A" in Figure 1A) was reproducibly accompanied by two additional minor peaks of activity (B and B' in Figure 1A). These activities appeared to be specific for the HFGAN72 receptor, since the active fractions failed to evoke detectable $[Ca^{2+}]_i$ transients in cells expressing an unrelated orphan receptor, HPRAJ70, in a parallel assay (Figure 1A). These activities seemed to be peptidic, since protease treatment of the active fractions destroyed the activity.

We purified the major activity to homogeneity from acid-acetone extracts of rat brains by collecting active fractions after four steps of HPLC (Figure 1). The final purified material, corresponding to the absorbance peak at the arrow in Figure 1D, was subjected to structural analyses by limited protease digestions followed by nanoelectrospray tandem mass spectrometry, Edman degradation, and MALDI mass spectrometry (see Experimental Procedures for details). The active substance, termed orexin-A, was found to be a 33-amino acid peptide of 3,562 Da, with an N-terminal pyroglutamyl residue and C-terminal amidation (Figure 2A). Molecular mass of the peptide as well as its sequencing analyses indicated that the four Cys residues of orexin-A formed two sets of

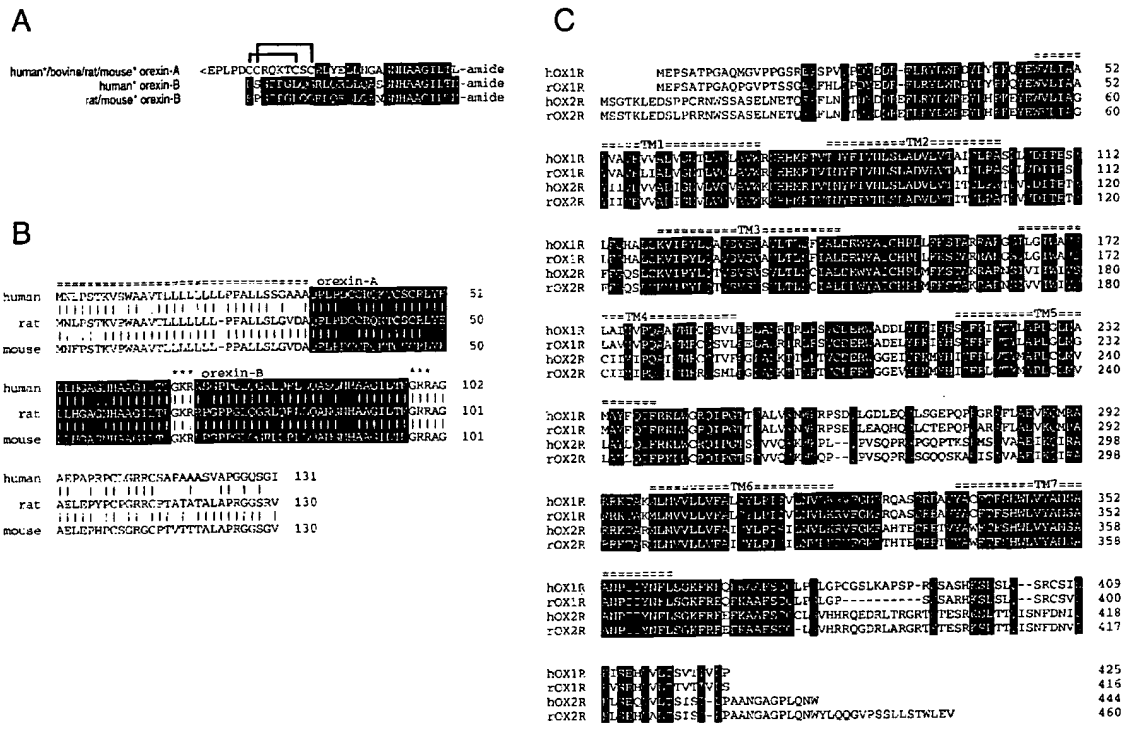


Figure 2. Amino Acid Sequences of Orexins and Orexin Receptors

(A) Structures of mature orexin-A and -B peptides. The topology of the two intrachain disulfide bonds in orexin-A is indicated above the sequence. The N-terminal pyroglutamyl residue (<E) in orexin-A is depicted. Amino acid identities are indicated by boxes. Asterisks indicate that human and mouse sequences are deduced from the respective genomic sequences and not from purified peptides.

(B) Deduced amino acid sequences of prepro-orexin precursor polypeptides. Orexin-A and -B sequences are boxed. Secretory signal sequences predicated by the SignalP Server (http://www.cbs.dtu.dk/services/SignalP/) are marked by equal signs (=); prohormone convertase cleavage and amidation sites by asterisks (*). Interspecifically identical residues are indicated by vertical bars. The full-length nucleotide sequences of human, rat, and mouse prepro-orexin cDNAs have been submitted to GenBank (accession numbers AF041240, AF041241, and AF041242).

(C) Deduced amino acid sequences of rat (r) and human (h) OX_1 and OX_2 receptors. Amino acid residues that are identical in all four sequences are boxed. Putative transmembrane (TM) domains are marked above the aligned sequences, as predicted by the PredictProtein server (http://www.embl-heidelberg.de/predictprotein). Gaps introduced to obtain optimal alignment are indicated by dashes. The full-length nucleotide sequences of human and rat orexin receptor cDNAs have been submitted to GenBank (accession numbers AF041243, AF041244, AF041245, and AF041246).

intrachain disulfide bonds. The topology of the disulfide bonds was determined by selectively synthesizing the peptides with different disulfide topologies by using selective deprotection and sequential oxidation of two of the four Cys residues and then comparing these synthetic peptides with purified natural orexin-A in isocratic reverse-phase HPLC. The topology was determined to be [Cys6-Cys12, Cys7-Cys14] (Figure 2A). Using similar protocols, we also purified the analogous activity from bovine hypothalamic extracts, and we found that the structure of bovine orexin-A was identical to that of the rat peptide.

In addition to the major peak of activity, the HPLC fractions contained two minor activity peaks that we designated B and B′ (Figure 1A). We isolated peptides responsible for both of these activities from rat brain extracts using similar purification protocols (see Experimental Procedures). Mass spectrometry and Edman sequencing showed that peak B contained a 28-amino acid, C-terminally amidated linear peptide of 2,937 Da. The peptide, termed orexin-B, was 46% (13/28) identical in sequence to orexin-A (Figure 2A). Peak B′ contained an N-terminally truncated [des Arg-Pro]orexin-B, which we designate orexin-B[3–28]. Further investigation is required to determine whether this truncation represents a physiologically relevant processing or whether it is a result of an artifactual proteolysis during the extraction procedure. In the subsequent studies described below, only orexin-A and full-length orexin-B were subjected to further biological characterization.

Structure of the Prepro-Orexin Precursor

The amino acid sequences of purified orexin-A and -B exhibited no meaningful similarity to any known peptides. A BLAST search of the GenBank database also failed to detect an entry that contains these or similar sequences. Therefore, we sought to isolate the cDNA encoding the precursor polypeptide; we first obtained a cDNA fragment encoding a part of orexin-A by RT–PCR of rat brain mRNA, using highly degenerate primers

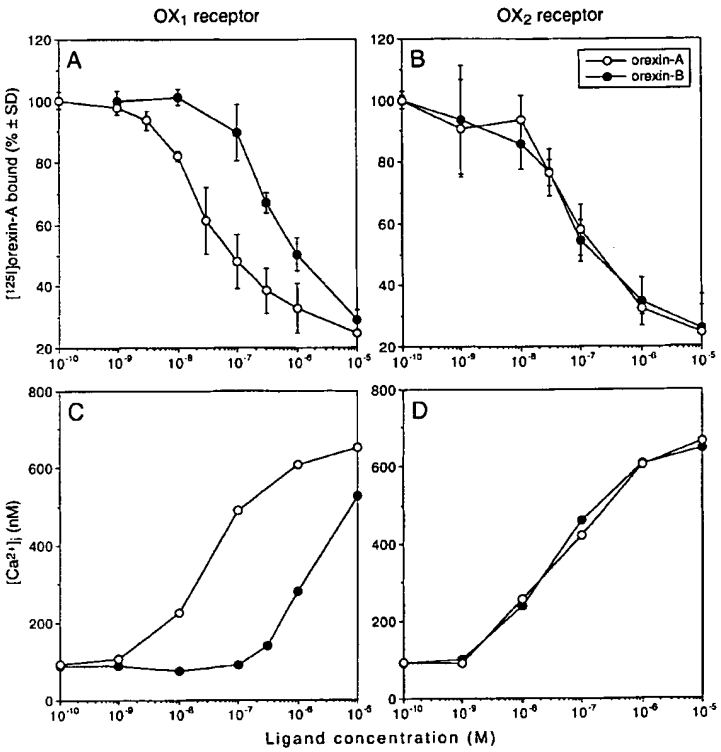


Figure 3. Pharmacological Characterization of Synthetic Human Orexins on Human Orexin Receptors Expressed in Stably Transfected CHO Cells

(A and B) Competitive radioligand binding assays. Displacement of [^{125}I-Tyr17]orexin-A binding to cells expressing human *OX$_1$R* (A) and *OX$_2$R* (B) by increasing concentrations of unlabeled orexin-A (open circle) and human orexin-B (filled circle) are determined in quadruplicates. Level of nonspecific binding was approximately 20% of the binding in the absence of competitor.

(C and D) Dose-response relationships of $[Ca^{2+}]_i$ transients evoked by orexin-A (open circle) and human orexin-B (filled circle) in cells expressing human *OX$_1$R* (C) and *OX$_2$R* (D). Peak $[Ca^{2+}]_i$ values are determined in duplicates.

based on parts of the orexin-A sequence. We then performed 5′-RACE and 3′-RACE reactions to obtain the full-length cDNA. We found that this cDNA encoded both orexin-A and -B (Figures 2A and 2B).

The 5′-most ATG codon of the 585 bp cDNA (GenBank accession number AF041241) was preceded by an in-frame stop codon, and the sequence around this initiation codon conformed to Kozak's rules. The open reading frame starting with this ATG encodes a 130-residue polypeptide, the rat prepro-orexin (Figure 2B). The first 33 amino acids exhibited characteristics of a secretory signal sequence: a hydrophobic core followed by residues with small polar side chains (von Heijne, 1986). The SignalP Server web site predicted that Ala32-Gln33 was the most likely site for signal sequence cleavage. The orexin-A sequence starts with Gln33, which is presumably cyclized enzymatically into the N-terminal pyroglutamyl residue by transamidation (Busby et al., 1987; Bateman et al., 1990). Thus, the mature peptide directly follows the signal peptide cleavage site. The last residue of the mature peptide is followed by Gly66, which presumably serves as NH_2 donor for C-terminal amidation by sequential actions of peptidylglycine monooxygenase and peptidylamidoglycolate lyase (Bradbury and Smyth, 1991; Eipper et al., 1993). As expected, Gly66 is followed by a pair of basic amino acid residues, Lys67-Arg68, which constitute a recognition site for prohormone convertases (Rouille et al., 1995). The next segment of the deduced prepro-orexin sequence, Arg69-Met96, was identical to the sequence of purified orexin-B. The Met96 residue is again followed by Gly-Arg-Arg, a C-terminal amidation signal, consistent with our finding that orexin-B is also C-terminally amidated.

We isolated human and mouse genomic fragments containing the *prepro-orexin* gene in order to predict the sequences of the human and mouse prepro-orexin polypeptides. Details of the gene structure will be reported elsewhere. The predicted human and mouse orexin-A sequences were both identical to rat and bovine orexin-A (Figure 2A). Mouse orexin-B was also predicted to be identical to rat orexin-B. Human orexin-B had two amino acid substitutions when compared with the rodent sequence (Figure 2A). Overall, the human and mouse prepro-orexin sequences were 83% and 95% identical to their rat counterparts, respectively. A majority of amino acid substitutions were found in the C-terminal part of the precursor, which appears unlikely to encode for another bioactive peptide.

Radiation hybrid mapping showed that the human *prepro-orexin* gene is most tightly linked to the MIT STS markers WI-6595 and UTR9641. The inferred cytogenetic location between these markers is 17q21. Interestingly, the localization at chromosome 17q21 raises the possibility that the prepro-orexin gene may be a candidate gene for a group of neurodegenerative disorders collectively called "chromosome 17-linked dementia" (Wilhelmsen, 1997), including nosological entities such as disinhibition-dementia-parkinsonism-amyotrophy complex (*DDPAC*; MIM No. *600274) and pallido-pontonigral degeneration (*PPND*; MIM No. *168610), which

may be caused by allelic mutations. Both *DDPAC* and *PPND* have recently been mapped to 17q21–22 (Wilhelmsen et al., 1994; Wijker et al., 1996).

Characterization of Orexin Receptors

Figure 2C shows the deduced amino acid sequences of the original HFGAN72 receptor, which we now call OX_1 receptor (OX_1R). Among various classes of GPCR, OX_1R is structurally most similar to certain neuropeptide receptors, most notably to the Y2 neuropeptide Y (NPY) receptor (26% identity), followed by the TRH receptor, cholesystokinin type-A receptor, and NK2 neurokinin receptor (25%, 23%, and 20% identity, respectively). This is consistent with our hypothesis that OX_1R is the receptor for orexins, another class of small regulatory peptides. In order to characterize further their pharmacological interactions, we performed in vitro functional assays using transfected cell lines expressing the receptor. Mock transfected CHO cells did not exhibit detectable levels of specific binding of radio-iodinated [^{125}I-Tyr17]orexin-A. Stable transfection of CHO cells with an expression vector containing the human *OX_1R* cDNA (CHO/OX_1R) conferred the ability to bind [^{125}I]orexin-A (Figure 3A). The radioligand binding was inhibited by nanomolar concentrations of unlabeled synthetic orexin-A in a dose-dependent manner, but not by any of several unrelated peptides tested, including human NPY and endothelin-1 at up to 10 μM (data not shown). The concentration of unlabeled orexin-A required to displace 50% of specific radioligand binding (IC_{50}) was 20 nM as calculated by the LIGAND program (Munson and Rodbard, 1980). Orexin-A also induced a transient increase in $[Ca^{2+}]_i$ in CHO/OX_1R cells in a dose-dependent manner (Figure 3C), but failed to induce detectable $[Ca^{2+}]_i$ transients in mock transfected CHO cells. We feel that the calcium mobilization is likely caused by the activation of the Gq class of heterotrimeric G proteins (Hepler et al., 1994). The calculated concentration of orexin-A required to induce half-maximum response (EC_{50}) was 30 nM. We obtained similar results in radioligand binding and $[Ca^{2+}]_i$ transient assays performed with stably transfected HEK293 cells (data not shown). These findings confirm that orexin-A is indeed a specific, high-affinity agonist for OX_1R.

As expected from our purification experiments, synthetic human orexin-B also acted as a specific agonist on CHO/OX_1R cells in a parallel set of experiments (Figures 3A and 3C). Interestingly, however, we found that human orexin-B has significantly lower affinity for the human OX_1R: the calculated IC_{50} in the competitive binding assay and the EC_{50} in the $[Ca^{2+}]_i$ transient assay were 420 nM and 2500 nM for human orexin-B, respectively, indicating 2–3 orders of magnitude lower affinities as compared with orexin-A. Similar results were obtained with HEK293 cells expressing *OX_1R*.

The findings with orexin-B led us to suspect that there is an additional orexin receptor(s) for which orexin-B has high affinity. A BLAST (tblastn) search of the GenBank dbEST database with the OX_1R amino acid sequence detected two highly similar EST entries: accession numbers D81887 (human fetal brain cDNA) and W86548 (human fetal liver/spleen cDNA). These EST sequences

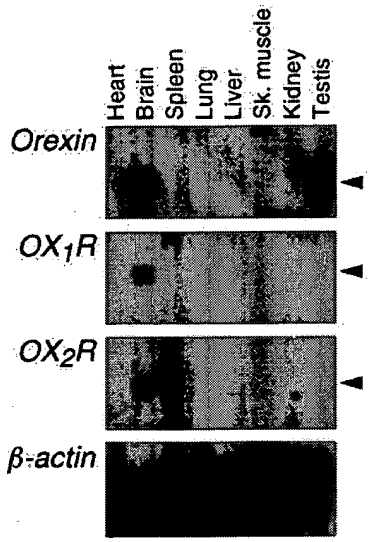


Figure 4. Tissue Distribution of Rat *prepro-orexin* and Orexin Receptor mRNAs

A membrane containing ~2 μg/lane of poly(A)$^+$ RNA from indicated rat tissues (Clontech, cat. no. 7764-1) was sequentially rehybridized with designated cDNA probes using standard methods. Sizes of *prepro-orexin*, *OX_1R* and *OX_2R* mRNAs are approximately 0.7, 2.5, and 3.5 kb, respectively (arrowheads). A segment of rat *prepro-orexin* cDNA encoding Gln33-Ser128 (Figure 2B) as well as full-length rat *OX_1R* and *OX_2R* cDNAs (Figure 2C) were used as probes. These probes do not detectably cross-hybridize with each other under the high-stringency conditions used. Autoradiography was exposed for 5 days, 3 days, 3 days and 3 hr for *prepro-orexin*, *OX_1R*, *OX_2R*, and *β-actin* probes.

showed much higher similarities to the OX_1R sequence than to numerous other GPCR entries. At the nucleotide sequence level, they differed enough from the human *OX_1R* cDNA sequence to indicate that they are not derived from the *OX_1R* gene itself. We subsequently found that these two ESTs were derived from 5′ and 3′ parts of the same transcript. Full-length cloning and sequencing showed that this cDNA encodes a GPCR, termed OX_2R, that highly resembles OX_1R (Figure 2C). The amino acid identity between the deduced full-length human OX_1R and OX_2R sequences is 64%. Thus, these receptors are much more similar to each other than they are to other GPCRs. Amino acid identities between the human and rat homologs of each of these receptors are 94% for OX_1R and 95% for OX_2R, indicating that both receptor genes are highly conserved between the species.

To characterize functionally the OX_2R further, we repeated competitive radioligand binding assays and $[Ca^{2+}]_i$ transient dose-response studies using stably transfected CHO cells expressing the human *OX_2R* cDNA. The results demonstrated that OX_2R is indeed a high-affinity receptor for human orexin-B, with an IC_{50} of 36 nM in the binding assay and an EC_{50} of 60 nM in the $[Ca^{2+}]_i$ transient assay (Figures 3B and 3D). Orexin-A also had high affinity for this receptor, with 38 nM IC_{50} and 34 nM EC_{50} values, which are similar to the values for orexin-B. Thus, we conclude that OX_2R is a nonselective

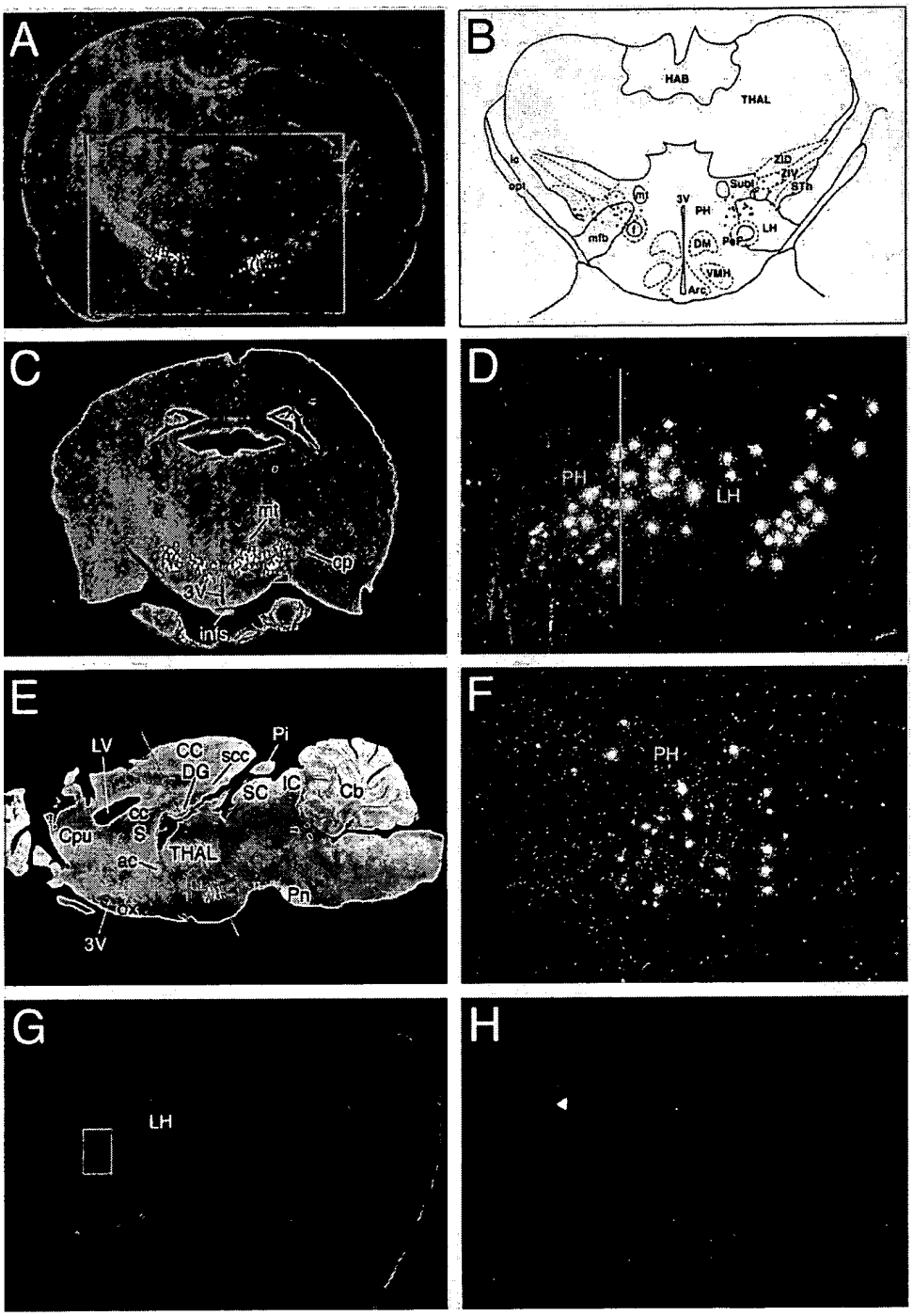


Figure 5. Localization of *prepro-orexin* mRNA and Immunoreactive Peptide in Adult Rat Brain

(A) Visualization of neurons containing *prepro-orexin* mRNA in the adult rat hypo- and subthalamic areas by in situ hybridization. Coronal section of brain hybridized with ^{35}S-labeled anti-sense riboprobe showing the bilateral and symmetrical distribution of labeled neurons. No detectable signal beyond background was generated by sense riboprobe. The yellow rectangle designates the area schematized in Figure 5B.

(B) *prepro-orexin* mRNA–containing neurons are shown in red superimposed upon anatomical structures of the hypo- and subthalamic areas (Paxinos and Watson, 1986): lateral hypothalamic area (LH), perifornical nucleus (PeF), posterior hypothalamic area (PH), subthalamic nucleus (Sth), subincertal nucleus (Subl), and ventral zona incerta (ZIV). Additional landmarks include: thalamus (THAL), habenular complex (HAB), internal capsule (ic), optic tract (opt), mammillothalamic tract (mt), fornix (f), medial forebrain bundle (mfb), third ventricle (3V), arcuate hypothalamic nucleus (Arc), dorsomedial hypothalamic nucleus (DM), and ventromedial hypothalamic nucleus (VMH). (Permission for use granted by Academic Press.)

receptor for both orexin-A and -B, while OX_1R is selective for orexin-A.

In radiation hybrid mapping, the MIT markers showing tightest linkage to the human OX_1R and OX_2R genes are the STS markers D1S195 and D1S443, and WI-5448 and CHLC.GATA74F07, respectively. The inferred cytogenetic locations between these markers are 1p33 for OX_1R, and 6cen (p11-q11) for OX_2R (accurate cytogenetic locations are often difficult to interpret from radiation hybrid maps in which the gene lies near the centromere).

Tissue Distribution of Orexin and Orexin Receptors

N-terminal pyroglutamyl cyclization (seen in orexin-A) as well as C-terminal amidation (both orexin-A and -B) are posttranslational modifications that are most often found in neuropeptides. Northern blot analysis of adult rat tissues showed that the 0.7 kb *prepro-orexin* mRNA is detected exclusively in the brain except, in a small amount, in the testis (Figure 4). In adult rats, OX_1R and OX_2R mRNAs were also detected exclusively in the brain (Figure 4). These findings are consistent with the idea that orexins are regulatory peptides that function within the central nervous system.

Preliminary Northern blot analysis of human mRNA from multiple brain regions (human brain blots II and III, Clontech) demonstrated that the *prepro-orexin* mRNA was expressed abundantly in subthalamic nucleus, but undetectable in other brain regions tested (data not shown; hypothalamus was not included in these brain blots). *prepro-orexin* mRNA was also undetectable in the human heart, placenta, lung, liver, skeletal muscle, kidney, and pancreas (data not shown). These findings are consistent with the results of the in situ hybridization in the rat brain described below.

Distribution of Orexin-Containing Neurons in Rat Brain

To further localize orexin expression within the central nervous system, we performed in situ hybridization and immunohistochemical analyses in rat brains. In situ hybridization with a rat *prepro-orexin* cRNA probe in coronal sections showed orexin-containing neurons organized bilaterally and symmetrically in a discrete set in hypothalamic and subthalamic areas of the adult rat brain (Figures 5A–5D). In the hypothalamus, positive neurons were found in the lateral (Figures 5A–5H) and posterior hypothalamic areas (Figures 5A–5F) and the perifornical nucleus (Figures 5C–5F). In the subthalamus (also called ventral thalamus), the zona incerta, subincertal, and subthalamic nuclei were positive (Figure 5B). Both in situ hybridization histochemistry (Figures 5A and 5C–5F) and immunofluorescent cytochemistry labeled the cytoplasm of medium-sized neurons (Figures 5D–5H), showing identical distributions. The shapes of the neurons varied from thin and fusiform to robust and multipolar. It is noteworthy that no signal was detected in neurons of the paraventricular (not shown), ventromedial (Figure 5B), or arcuate nuclei (Figure 5B), which are known to contain a variety of neuropeptides associated with food consumption (Bernardis and Bellinger, 1996; Bing et al., 1996; see below).

Stimulation of Feeding Behavior by Centrally Administered Orexins

The lateral hypothalamic area has been implicated in the regulation of feeding behavior and energy homeostasis ever since the classic experiments showing that animals with lateral hypothalamic lesions had decreased food intake and lower "set point" for body weight (Oomura, 1980; Bernardis and Bellinger, 1993, 1996). The striking localization of orexin–containing neurons in the lateral hypothalamus and some of its adjacent areas suggested to us the possibility that the neuropeptide may be involved in the regulation of food intake. To evaluate this hypothesis, we administered orexin acutely into the lateral ventricle of male rats through preimplanted indwelling catheters. When a bolus was administered intracerebroventricularly in early light phase, orexin-A stimulated food consumption in a dose-dependent manner within 1 hr (Figure 6, left). The magnitudes of stimulation with 3 nmol and 30 nmol orexin-A the 2 hr time point were 6- and 10-fold, respectively. The effect persisted at 4 h; the amount of food consumed during the interval from 2–4 h postinjection was increased approximately 3-fold with either dose as compared to vehicle controls. Human orexin-B also significantly augmented food intake; at the 2 hr time point, we observed 5- and 12-fold stimulation of food consumption by 3 nmol and 30 nmol orexin-B, respectively, as compared with vehicle controls (Figure

(C) Low magnification of section caudal to Figure 5A showing the distribution of *prepro-orexin* mRNA–containing neurons. Mammillothalamic tract (mt), third ventricle (3V), cerebral peduncle (cp), infundibular stem (Infs). There are two groups of neurons: one group is seen in the posterior hypothalamus (PH) and the other in the lateral hypothalamus (LH). The yellow rectagle indicates the field shown in Figure 5D.

(D) Higher magnification of the neuronal cell bodies in Figure 5C. The yellow vertical line indicates the boundary separating the medial neurons of PH and the neurons of LH.

(E) Low magnification of *prepro-orexin* mRNA–containing neurons seen in parasagittal section traversing the posterior hypothalamic area shown in Figure 5D. Cerebral cortex (CC), dentate gyrus (DG), splenium of the corpus callosum (scc), lateral ventricle (LV), caudate putamen (Cpu), corpus callosum (cc), septum (S), anterior commissure (ac), thalamus (THAL), optic chiasm (ox), third ventricle (3V), pineal gland (Pi), superior colliculus (SC), inferior colliculus (IC), cerebellum (Cb), and pontine nucleus (Pn). The yellow rectagle indicates the field magnified in Figure 5F.

(F) High-magnification of the *prepro-orexin*–expressing neurons in the posterior hypothalamus (PH) shown in Figure 5E.

(G) Orexin-A-immunoreactive neurons in the LH of a frozen parasagittal section. Dorsal aspect is to the left side of this panel. The yellow box corresponds to the field shown in Figure 5H. The staining of orexin-A-positive neurons was abolished in the presence of excess (100 nM) synthetic orexin-A.

(H) High magnification of the immunoreactive orexin–containing neurons of the LH area. The boxed area of cells Figure 5G have been rotated 90° clockwise. In general, most of the neurons are multipolar, but a few are fusiform. The nucleus is not immunopositive, indicating that orexin is cytoplasmic. Beaded neuronal processes are sometimes seen in these sections (arrowhead).

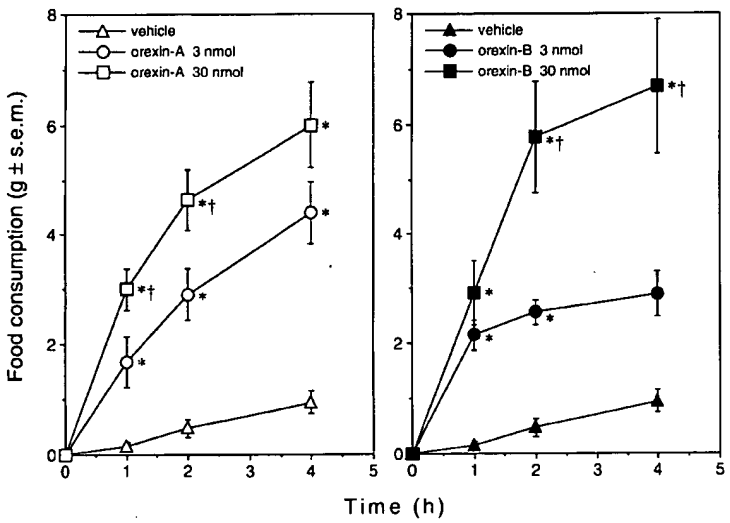


Figure 6. Stimulation of Food Consumption by Intracerebroventricular Injection of Orexin-A and -B in Freely Fed Rats

Designated amounts of synthetic human orexin-A (left) or -B (right) were administered in a 5 μl bolus through a catheter placed in the left lateral ventricle in early light phase. Cumulative food consumption was plotted over the period of 4 hr after injection. Asterisks (*) indicate significant difference from vehicle controls ($p < 0.05$, n = 8–10, ANOVA followed by Student-Newman-Keuls test). Crosses (†) designate significant difference between 3 nmol and 30 nmol injections ($p < 0.05$, n = 8–10, ANOVA followed by Student-Newman-Keuls test). Similar results were obtained in at least four independent sets of experiments. The same vehicle control curve was replotted in both panels.

6, right). The effects of orexin-B did not last as long as those of orexin-A; there was little stimulation of food intake after 2 hr even with the higher dose. These actions of orexin were less efficacious than NPY, a known orexigenic neuropeptide (Gerald et al., 1996; Turton et al., 1996), but longer lasting in the case of orexin-A; under the same conditions, we found that 3 nmol human NPY induced cumulative food consumption of 11.4 ± 2.0 g and 12.8 ± 1.8 g (mean ± SEM) at 2 hr and 4 hr internals, respectively, representing 24- and 14-fold stimulation over vehicle controls.

The shorter action of orexin-B as compared with orexin-A might be related to the fact that orexin-B is a linear peptide with a free N terminus. In contrast, both termini of orexin-A are blocked by posttranslational modifications. Together with the two intrachain disulfide bonds, these modifications may render orexin-A resistant to inactivating peptidases. We cannot exclude the possibility that these time courses reflect bona fide differences in the biological actions of orexin-A and orexin-B in the central nervous system. The similar degrees of stimulation of food intake by both orexin-A and orexin-B suggest that OX_2R, the nonselective receptor (see Figure 3), may at least in part be involved in this in vivo pharmacological action of orexins. However, more detailed in vivo dose-response studies, as well as a separate blockade of these receptors either pharmacologically by selective antagonists or genetically by gene targeting, will be required to determine the receptor(s) involved.

Up-Regulation of *prepro-orexin* mRNA in the Fasting State

The stimulation of food intake by centrally administered orexins suggested that these neuropeptides may play a physiological role in the central regulation of feeding. To evaluate the possibility that orexin production and release may be affected by nutritional state, we compared the level of *prepro-orexin* mRNA expression in the hypothalamus of fed and fasted rats. Adult male rats were housed singly under 12 hr light-dark cycle for at least 5 days, and then either fed ad libitum (n = 5) or fasted for 48 hr starting at mid-light cycle (n = 5). At the end of the 48 hr period, rats were sacrificed and the thalamic/hypothalamic portion of the brain was dissected. Total RNA was extracted from individual samples and subjected to quantitative Northern blot analysis. After the 48 hr fast, hypothalamic *prepro-orexin* mRNA was up-regulated 2.4-fold as compared with the fed control animals (Figure 7). *NPY* mRNA was previously described to be up-regulated under similar fasted conditions (Qu et al., 1996). We also found that *NPY* mRNA was up-regulated, albeit to a lesser extent than the orexin mRNA (Figure 7A).

Physiological Implications

Quantum leaps in our understanding of the mechanisms involved in the energy homeostasis have resulted from the recent identification of the leptin signaling pathway (Zhang et al., 1994; Chen et al., 1996) and by additional molecular genetic studies in mice. These studies have revealed a number of central regulatory pathways mediated by various neuropeptides (Yen et al., 1994; Erickson et al., 1996a; Huszar et al., 1997; Ohki-Hamazaki et al., 1997). Maintenance of body weight is achieved by an intricate balance between energy intake (food consumption) and expenditure. This energy homeostasis is ultimately governed by the brain, where a variety of afferent signals reflect the animal's nutritional state and its external environment, particularly food availability. These signals are integrated in order to moderate efferent pathways that control feeding behavior and energy expenditure (Bray, 1995; Rosenbaum et al., 1997). The afferent signals include somatic sensory signals like smell and taste, gastrointestinal signals like mechanical distention and chemical stimuli to the mucosa, blood-born metabolites like glucose and free fatty acids, and "lipostatic" signals such as leptin (Friedman, 1997). Together with the regulation of metabolic rate, which is in part mediated by sympathetic and parasympathetic nervous systems and thyroid hormone, the regulation of feeding behavior constitutes the efferent wing of energy homeostasis. The central site of integration is the hypothalamus (Oomura, 1980; Bernardis and Bellinger, 1993, 1996),

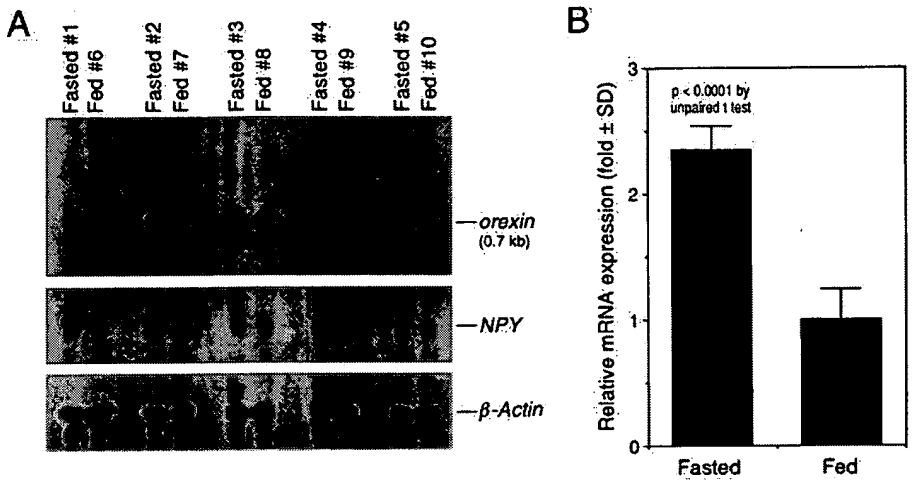


Figure 7. Up-Regulation of *prepro-orexin* mRNA in Fasted Rats

(A) Northern blot analysis of hypothalamic *prepro-orexin* mRNA expression in freely fed and 48 hr fasted rats. Each lane contains 20 μg of total RNA from rat diencephalon. The membrane was rehybridized with *NPY* cDNA and then with β-*actin* cDNA as internal reference for the amount of RNA loaded.

(B) Densitometric quantitation of Northern blot results presented in Figure 7A. The ratio of *prepro-orexin* and β-*actin* mRNA signals was determined in each lane, and these normalized values were then compared between the fasted and fed groups. The mean value for normalized *prepro-orexin* mRNA expression in the fed group was arbitrarily designated as unity.

where a large array of neurotransmitters, especially neuropeptides, modulate signals through complex neural circuits.

Recent molecular genetic studies in rodents have started to provide convincing evidence supporting important roles for certain hypothalamic neuropeptides in the central regulation of energy balance (Wolf, 1997). For example, NPY, which is negatively regulated by leptin (Zarjevski et al., 1993; Stephens et al., 1995; Tartaglia et al., 1995), has been established as one of the positive regulators of feeding behavior, presumably acting on the Y5 NPY receptor (Gerald et al., 1996); thus, the obesity of leptin-deficient mice is partially ameliorated when they are also deficient in NPY (Erickson et al., 1996b). Ectopic and constitutive expression of the endogenous antagonist for melanocortin receptors, the Agouti protein, is the basis of obesity in obese-yellow mice (Yen et al., 1994; Fan et al., 1997), and, indeed, deficiency in the MC4 melanocortin receptor completely replicates the obesity syndrome of obese-yellow mice (Huszar et al., 1997). Furthermore, Agouti-related protein (also termed ART), another endogenous antagonist for the MC4 receptor, has recently been shown to be expressed in the hypothalamus, and the expression is up-regulated in *ob/ob* and *db/db* mice (Ollmann et al., 1997; Shutter et al., 1997). More recently, mice deficient for an orphan neuropeptide receptor (bombesin receptor type-3) have also been shown to become obese (Ohki-Hamazaki et al., 1997). Within the hypothalamus, two regions have traditionally been associated with the regulation of feeding and energy balance: the ventromedial hypothalamus and the lateral hypothalamic area. A classical series of lesion experiments led to the hypothesis that the ventromedial hypothalamus functions as a "satiety center," whereas the lateral hypothalamic area works as a "feeding center." Thus, a lateral hypothalamic lesion causes the animal to eat less and lose weight, sometimes leading to death by starvation. While clearly oversimplified, these hypotheses still provide a valid conceptual framework (Oomura, 1980).

Recent studies continue to reveal the molecular basis for the role of periventricular/medial hypothalamic regions in energy homeostasis, e.g., ventromedial nucleus, arcuate nucleus, and paraventricular nucleus. Neurons containing neuropeptides such as NPY (Bing et al., 1996), melanocortins (Jacobowitz and O'Donohue, 1978), glucagon-like peptide-1 (Shughrue et al., 1996), and galanin (Warembourg and Jolivet, 1993), as well as the leptin and melanocortin-4 receptors (Mountjoy et al., 1994; Tartaglia et al., 1995) are abundant in one or more of these periventricular/medial hypothalamic regions. In contrast, few neuropeptides have been described to be produced chiefly in the lateral hypothalamic regions. Other than the orexins, we are aware of only one distinct neurotransmitter that is specifically produced in the lateral hypothalamus: melanin concentrating hormone (MCH) has been localized in the zona incerta and the lateral hypothalamic area (Bittencourt et al., 1992). Intriguingly, MCH was recently reported to stimulate food intake upon central administration. Moreover, *MCH* mRNA is up-regulated in *ob/ob* mice and by fasting in wild-type mice (Qu et al., 1996). It will be important to investigate further the possible interplay of orexin with this and other positive (e.g., Agouti-related protein, NPY, galanin, and opioids) and negative (e.g., leptin, melanocortins, corticotropin-releasing factor, glucagon-like peptide-1, and cholecystokinin) regulators of energy balance both within and outside the central nervous system (Arase et al., 1988; Rosenbaum et al., 1997).

A decline of blood glucose levels can signal the initiation of food intake (Oomura, 1980). The lateral hypothalamic area contains glucose-sensitive neurons that are

activated by glucopenia and thus implicated in the positive short-term regulation of feeding and energy expenditure (Oomura et al., 1974). It is tempting to speculate that all or some of the orexin-containing neurons may be glucose-sensitive, or that they may receive stimulatory projections from glucose-sensitive neurons. Future experiments will also determine whether orexins have additional actions relevant to nutritional homeostasis, such as effects on the regulation of systemic energy expenditure, secretion of metabolic hormones such as insulin, and ultimately, the regulation of body weight.

The present discovery of orexins and their receptors may provide a novel molecular basis for the role of the lateral hypothalamic areas in the regulation of feeding behavior. It is clear, however, that the definitive assignment of physiological roles for orexins requires further pharmacological as well as molecular genetic investigations. Nevertheless, pharmacological intervention directed at the orexin receptors may prove to be an attractive avenue toward the discovery of novel therapeutics for diseases involving disregulation of energy homeostasis, such as obesity and diabetes mellitus.

Experimental Procedures

Transfected Cells

Full-length cDNAs encoding individual orphan GPCRs were subcloned into the pCDN mammalian expression vector. These plasmids were individually transfected into the human embryonic kidney (HEK) 293 cells, and stable transfectant cell lines established and maintained as described (Aiyar et al., 1994). For in vitro pharmacological assessments (Figure 3), CHO-K1 cells were stably transfected with human *OX_1R* and *OX_2R* expression vectors, and transfected cell lines were maintained as previously described (Xu et al., 1994).

Purification of Orexin

Approximately 200 g of frozen rat whole brain (150 pieces, Pel-Freeze) were homogenized by Polytron in 10 × volume of 70% (v/v) acetone, 1 M acetic acid, and 20 mM HCl. The homogenate was centrifuged at 20,000 × g for 30 min at 4°C. The resultant supernatant was collected and extracted three times with diethyl ether. The aqueous phase was centrifuged at 20,000 × g for 30 min at 4°C, and the supernatant was loaded onto two 10 g cartridges of Sep-Pak C18 (Waters), which were preequilibrated with 0.1% (v/v) TFA. Cartridges were washed with 5% CH_3CN/0.1% TFA, and then eluted with 50% CH_3CN/0.1% TFA. The eluate was lyophilized, redissolved in 1 M acetic acid with sonication, and then filtered through a 20 μm Mirex GV filter (Millipore). Step 1 (Figure 1A): the extract was directly loaded onto a C_{18} reverse-phase HPLC column (Vydac 218TP510; 10 mm × 250 mm), preequilibrated with 3% CH_3CN/0.1% TFA at a flow rate of 3 ml/min at room temperature. A 10%–40% gradient of CH_3CN in 0.1% TFA was then applied over 100 min. Three-milliliter fractions were collected, and ≈2% of each fraction was set aside and assayed for $[Ca^{2+}]_i$ transients in OX_1R–expressing HEK293 cells (see below). Step 2 (Figure 1B): the active fractions were pooled and directly applied to a cation-exchange HPLC column (TosoHaas SP-5PW; 7.5 mm × 75 mm) preequilibrated with 20 mM Na-phosphate (pH 3.0)/30% CH_3CN at room temperature. A 0–0.5 M gradient of NaCl in 20 mM Na-phosphate (pH 3.0)/30% CH_3CN was applied over 60 min at a flow rate of 1 ml/min. One-milliliter fractions were collected, and ~3% from each fraction was used for the $[Ca^{2+}]_i$ assay. Step 3 (Figure 1C): the active fractions were pooled, diluted 4-fold with 0.1% TFA, and directly loaded onto an analytical C_{18} reverse-phase column (Vydac 218TP54; 4.6 mm × 250 mm) preequilibrated with 3% CH_3CN/0.1% TFA at a flow rate of 1 ml/min at room temperature. A 21%–36% gradient of CH_3CN in 0.1% TFA was applied over 75 min. Individual UV absorption peaks (210 nm) were collected manually, and ~3% from each fraction was assayed. Step 4 (Figure 1D): the active peak was diluted 4-fold with 0.1% TFA and directly loaded onto the same C_{18} column, but this time preequilibrated at 40°C with 3% CH_3CN/0.1% TFA. A 21%-36% gradient of CH_3CN in 0.1% TFA was applied over 75 min at 40°C. The major 210 nm peak, representing virtually pure orexin-A, was collected manually. For purification of rat orexin-B and orexin-B[3–28] as well as bovine orexin-A, basically the same procedures were used, except that step 3 was replaced with a reverse-phase HPLC under neutral pH: active fractions from step 2 were directly loaded onto the same analytical C_{18} column preequilibrated with 3% CH_3CN/20 mM Tris-HCl (pH 7.0 at 40°C). A 3%–40% gradient of CH_3CN in 20 mM Tris-HCl (pH 7.0) was applied over 74 min at 40°C. Individual 210 nm peaks were collected manually, and the active fraction was subjected to step 4 as described above.

Peptide Microsequencing

Direct sequencing of the intact orexin-A peptide by Edman degradation failed, suggesting a blocked amino terminus. The monoisotopic molecular mass of the intact peptide was determined to be 3558.7 ± 0.1 Da by time lag–focusing matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (Carr and Annan, 1996) using internal peptide standards (Micromass TofSpec-SE). The measured mass of the peptide increased by 232 Da upon reaction with iodoacetamide in the presence of DTT, indicating the presence of two intramolecular disulfides bonds. Digestion of the reduced and alkylated peptide with Lys-C provided two peptides with average M_r = 1287 and 2530. Edman sequence of the mixture provided a single partial sequence of TCSCRLYELLHGAGNHAAG. Edman sequence analysis of an Asn-C digest mixture extended the C-terminal sequence as follows . . . HAAGILTL. The C-terminal sequence, . . . EXX HGAGNHAAGXXTX (where X is either Leu or Ile), was also confirmed by MALDI mass spectrometry of the peptide partially digested with carboxypeptidase Y (Voyager Elite XL, PerSeptive Biosystems). The blocked N-terminal Lys-C peptide was sequenced by nanoelectrospray collision induced dissociation (CID) tandem mass spectrometry (Carr et al., 1996; Wilm and Mann, 1996) on a triple quadrupole mass spectrometer (Perkin-Elmer Sciex API-III). A sequence of <EPXPDCCRQK (calculated MH^+ = 1286.6) was determined from the tandem MS data, where <E is pyroglutamic acid, and X is either Leu or Ile. The identity and sequence of the first two residues were confirmed by MS^3 of the b2 ion produced by in-source CID. X was determined to be Leu by amino acid analysis of the HPLC purified N-terminal peptide. The Gln residue at position 9 was distinguished from Lys (both amino acids have the same residue mass) by acetylation of the peptide and remeasurement of the molecular weight. The MH^+ ion for the ligand shifted by 42 Da from 1286.6 to 1328.6, indicating the addition of only one acetate group (calculated MH^+ = 1328.6). Because Gln residues cannot be acetylated and the N terminus is blocked, the addition of only one acetate group strongly suggests the C-terminal sequence is QK, not KK. The M, determined for the intact peptide (3558.7 ± 0.1 Da) indicates that the C terminus is an amide rather than a free acid and that all four Cys are in disulfide linkages (calculated monoisotopic M_r = 3558.7).

Cloning of Rat *prepro-orexin* cDNA

A part of the orexin-A sequence, QPLPDCCRQKTCSCRLYELLHGA GNHAAGI (amino acids 1–30), was chosen to design highly degenerate oligonucleotide primers encoding its ends: 5′-CA(A,G)CCN(C,T)T NCCNGA(C,T)TG(C,T)TG-3′ and 5′-ATNCCNGCNGC(A,G)TG(A,G)TT-3′. A cDNA fragment correctly encoding the peptide was obtained by RT–PCR using rat brain poly(A) RNA as template. A nondegenerate oligonucleotide primer, 5′-GTTGCCAGCTCCGTGCAACAGTTCGTA GAGACGG-3′, was designed based on the sequence of the RT–PCR product. Double-stranded cDNA was synthesized from rat brain poly(A) RNA, ligated to the Marathon adaptor (Clontech), and used as template for the initial 5′-RACE reaction with Adaptor Primer 1 (Clontech) and the above specific primer. A nested PCR reaction was performed using an oligonucleotide, 5′-CGGCAGGAACACGTC TTCTGGCG-3′, and the Adaptor Primer 2 (Clontech). An approximately 250 bp 5′ cDNA product, correctly encoding the peptide, was obtained. Single-stranded rat brain cDNA was then synthesized using an oligonucleotide, 5′-CCTCTGAAGGTTCCAGAATCGATAG $(T)_{25}$(A,C,G)N-3′, as a specific primer for the reverse transcription.

This was used as template for a 3′-RACE reaction using a specific primer, 5′-TCCTTGGGTATTTGGACCACTGCACCGAAG-3′ (corresponding to a part of the 5′-noncoding region of the cDNA sequence obtained by the 5′-RACE), and an anchor primer, 5′-CCTCTGAAGGT TCCAGAATCGATAG-3′. The product was subjected to a nested PCR reaction using an oligonucleotide, 5′-ATACCATCTCTCCGGAT TGCCTCTCCCTGA-3′, and the same anchor primer. A discrete 0.6 kb product containing the correct 5′ cDNA sequence was obtained. The full-length cDNA sequence was determined by resequencing the products obtained from three independent 3′-RACE reactions.

Cloning of Human and Mouse *prepro-orexin* Genes

Since we found that the full-length rat *prepro-orexin* cDNA, which contains CTG triplet repeats (encoding the oligo-leucine stretch in the signal sequence), tends to cross-hybridize with a number of unrelated genes, we produced a cDNA probe that did not contain the repeats by RT–PCR from rat brain RNA, using primers 5′-CAGCC TCTGCCCGACTGCTGT-3′ and 5′-CGTCTTTATTGCCTAGGGACTG GCGAGGAG-3′. Human and mouse genomic libraries (Clontech and Stratagene, respectively) were screened by plaque hybridization to this cDNA probe. Resulting genomic fragments containing the entire *prepro-orexin* coding sequences were further characterized and sequenced by standard procedures.

Cloning of *OX$_1$R* and *OX$_2$R* cDNAs

Cloning of full-length human *OX$_1$R* (HFGAN72) cDNA has been disclosed elsewhere (US Patent No. WO96/34877; November 7, 1996) (Figure 2C). Rat *OX$_1$R* cDNA was obtained by screening rat brain cDNA libraries with the full-length human *OX$_1$R* cDNA as probe, using standard procedures. In order to clone full-length human *OX$_2$R* cDNA, an inter-EST RT–PCR was performed with human brain poly(A)$^+$ RNA as template, using oligonucleotide primers based on the public 5′ and 3′ EST sequences (GenBank accession numbers D81887 and W86548, respectively). The resultant cDNA product, which contained a large part of the coding region spanning the transmembrane domains 1–7, was further extended by standard 5′-RACE and 3′-RACE procedures. Rat *OX$_2$R* cDNA was obtained by screening rat brain cDNA libraries with the full-length human *OX$_2$R* cDNA as probe.

Radiation Hybrid Mapping

The following oligonucleotides were used to amplify genomic fragments: *prepro-orexin*, GCCAAAGGTGTCTCCACTC and GACAGGT GCAAACGGAGCAC; *OX$_1$R*, GGTGCTCACCAGCGTCACCAC and GCTTAATCCTCACATCAACCCTGC; *OX$_2$R*, GAGGAGCTTGCAGCAT TGAGC and GTGCAGGTATTCCCTCCACAG. PCR of the *prepro-orexin*, *OX$_1$R*, and *OX$_2$R* genes was performed in duplicates against the 96 samples (including 3 controls: donor DNA, host DNA, and water) in the GeneBridge 4 radiation hybrid panel (Research Genetics) (Gyapay et al., 1996). The scoring strings of 93 × 1's (presence), 0's (absence), and 2's (ambiguous) were submitted to the Radiation Hybrid Mapping Server at MIT (http://www.genome.wi.mit.edu/) (Hudson et al., 1995). Cytogenetic locations were inferred from these results using the markers shown to neighbor the genes via The Genome Directory (1995), and GDB (http://gdbwww.gdb.org/gdb/docs/gdbhome.html). Chromosomal locations were confirmed by PCR on the BIOS panel (BIOS, USA) using the same oligo pairs and conditions.

Radioligand Binding Assay

Synthetic human orexin-A was ^{125}I-labeled at Tyr17 by Chloramine-T oxidation in the presence of Na^{125}I (2,000 Ci/mmol, New England Nuclear), and monoiodinated peptide was purified by C^{18} reverse-phase HPLC as described (Takigawa et al., 1995). Stable transfectant CHO cell lines expressing human *OX$_1$R* or *OX$_2$R* were each seeded onto 12-well plates at a density of 3 × 10^5 cells per well. After an overnight culture, medium was discarded and cells were incubated at 20°C for 90 min with binding buffer (HEPES-buffered saline/0.5% bovine serum albumin) containing 10^{-10} M [^{125}I]orexin-A plus designated concentrations of unlabeled competitors. Cells were then washed three times with ice-cold phosphate-buffered saline, lysed in 0.1 N NaOH, and cell-bound radioactivity was determined by a γ-counter.

Intracellular Calcium Transient Assay

Stable transfectant HEK293 cells expressing orphan GPCRs were loaded with Fura-2 AM in suspension, and $[Ca^{2+}]_i$ transients evoked by agonists were monitored by a Model CAF-110 Intracellular Ion Analyser (JASCO) in 500 μl cuvette as previously described (Grynkiewicz et al., 1985; Xu et al., 1994). For in vitro pharmacological characterization (Figure 3), the same procedures were performed using stably transfected CHO cells expressing human *OX$_1$R* or *OX$_2$R*.

In Situ Hybridization

A 0.29 kb segment of rat cDNA encoding Gln33-Ser128 of prepro-orexin was generated as described above and subcloned into pBluescript II SK(+) vector. Sense and anti-sense riboprobes were generated with T7 and T3 RNA polymerases, respectively, using the Maxiscript kit (Ambion) in the presence of ^{35}S-CTP (Amersham). In situ hybridization to adult rat brain sections was performed as described (Benjamin et al., 1997).

Immunohistochemistry

Anti-orexin-A antiserum was raised in rabbits by immunization with synthetic orexin-A[14–33], CRLYELLHGAGNHAAGILTL-amide, conjugated to keyhole limpet hemocyanin (Calbiochem) using 3-maleimidobensoic acid N-hydroxysuccimide ester (Sigma). Antiserum was affinity purified on an orexin-A-conjugated Sepharose CL-4B (Pharmacia) column and used for immunohistochemistry. Male Sprague-Dawley rats (~300 g) were anesthetized and perfused via the left cardiac ventricle with phosphate-buffered saline (PBS). Brain was taken out, directly embedded in OCT compound (Tissue-Tek), and frozen in liquid nitrogen. Cryostat sections (15 μm) were cut and mounted on silane-coated glass sides. The sections were post-fixed with 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.2) for 1 hr and washed three times in PBS. The sections were incubated with 1% bovine serum albumin in PBS for 1 hr and then incubated with affinity-purified antiserum in the same solution for 1 hr at room temperature. After washing three times in PBS, the sections were incubated with FITC-conjugated goat anti-rabbit IgG antibody (Cappel) for 1 hr at room temperature. Slides were then washed three times in PBS and examined under fluorescence microscope.

Intracerebroventricular Administration of Orexins

Male Wistar rats (180–200 g on arrival; Charles-River) were housed under controlled lighting (12 hr light-dark cycle) and temperature (22°C) conditions. Food (standard chow pellets) and water were available ad libitum. Rats (200–220 g) were anesthetized with pentobarbital (50 mg/kg i.p.), positioned in a Koph Model 900 stereotaxic frame, and implanted with a guide cannula into the left lateral ventricle under sterile conditions using a MEDIBIO Optical Brain Tracer (Muromachi Kikai) (Ikeda and Matsushita, 1980). The coordinates used to map the correct positioning of the implants were: 6.1 mm anterior to the lambda, 1.5 mm lateral from the midline, and ~3.4 mm (guided by MEDIBIO) ventral to the skull surface, with the incisor bar set 3.3 mm below the interauricular line. Rats were then housed singly under the same conditions as above for a recovery period of at least 7 days, and body weights were monitored daily for the duration of the study. After recovery from surgery, rats were transferred to grid-floor cages and fed with powdered chow so that food intake measurements could be made. The rats were acclimated to the new environment at least for 1 day. The position of the cannula was verified by central administration of human NPY (3 nmol in sterile water); for a positive test, at least 8 g of food was eaten over a 4 hr period postinjection. Only positively testing animals (n = 8–10) were used. The studies were conducted according to a multidose, crossover design, with the order of dosing determined using the Latin square principle, leaving at least one rest day between administrations. All doses were delivered in a volume of 5 μl in sterile water over 30 s, and the injector remained in position for a further 30 s to ensure complete dispersal of the peptide. All intracerebroventricular administrations began at 2 h into light cycle, and food

intake was measured at 1, 2, and 4 h intervals. All peptides were dissolved in sterile water, initially at 6 mM, and diluted in water as needed. Water alone was used for the vehicle control.

Acknowledgments

We thank Katsutoshi Goto for discussion, Tadahiro Nambu for technical assistance in immunohistochemistry, and Mike Brown and Joe Goldstein for critically reading the manuscript. We also thank the following scientists at SmithKline Beecham for their various contributions: Neil Upton, Christine Debouck, Catherine Ellis, Ganesh Sathe, Michael Huddleston, Jeff Stadel, John Martin, Lily Zhang, Paul Lysko, and Mary Brawner. M. Y. is an Investigator and T. S. and A. A. are Associates of the Howard Hughes Medical Institute. R. M. C. is an NIH fellow of the Pediatric Scientist Development Program. M. I. is a Summer Undergraduate Research Fellow of the University of Texas Southwestern Medical Center. This work was supported in part by research grants from the Perot Family Foundation.

Received December 12, 1997; revised January 7, 1998.

References

Aiyar, N., Baker, E., Wu, H.-L., Nambi, P., Edwards, R.M., Trill, J.J., Ellis, C., and Bergsma, D.J. (1994). Human AT_1 receptor is a single copy gene: characterization in a stable cell line. Mol. Cell Biochem. *131*, 75–86.

Arase, K., York, D.A., Shimizu, H., Shargill, N., and Bray, G.A. (1988). Effects of corticotropin-releasing factor on food intake and brown adipose tissue thermogenesis in rats. Am. J. Physiol. *255*, E255–259.

Bateman, A., Solomon, S., and Bennett, H.P.J. (1990). Post-translational modification of bovine pro-opiomelanocortin. J. Biol. Chem. *265*, 22130–22136.

Benjamin, I.J., Shelton, J., Garry, D.J., and Richardson, J.A. (1997). Temporospatial expression of the small HSP/αB-crystallin in cardiac and skeletal muscle during mouse development. Dev. Dyn. *208*, 75–84.

Bernardis, L.L., and Bellinger, L.L. (1993). The lateral hypothalamic area revisited: neuroanatomy, body weight regulation, neuroendocrinology and metabolism. Neurosci. Biobehav. Rev. *17*, 141–193.

Bernardis, L.L., and Bellinger, L.L. (1996). The lateral hypothalamic area revisited: ingestive behavior. Neurosci. Biobehav. Rev. *20*, 189–287.

Bing, C., Wang, Q., Pickavance, L., and Williams, G. (1996). The central regulation of energy homoeostasis: roles of neuropeptide Y and other brain peptides. Biochem. Soc. Trans. *24*, 559–565.

Bittencourt, J.C., Presse, F., Arias, C., Peto, C., Vaughan, J., Nahon, J.-L., Vale, W., and Sawchenko, P.E. (1992). The melanin-concentrating hormone system of the rat brain: an immuno- and hybridization histochemical characterization. J. Comp. Neurol. *319*, 218–245.

Bradbury, A.F., and Smyth, D.G. (1991). Peptide amidation. Trends Biochem. Sci. *16*, 112–115.

Bray, G.A. (1995). Obesity, fat intake, and chronic disease. In Psychopharmacology: The Fourth Generation of Progress, F.E. Bloom and D.J. Kupher, eds. (New York: Raven Press), pp. 1591–1608.

Busby, W.H., Jr., Quackenbush, G.E., Humm, J., Youngblood, W.W., and Kizer, J.S. (1987). An enzyme(s) that converts glutaminyl-peptides into pyroglutamyl-peptides. J. Biol. Chem. *262*, 8532–8536.

Carr, S.A., and Annan, R.S. (1996). Overview of peptide and protein analysis by mass spectrometry. In Current Protocols in Protein Science (New York: John Wiley and Sons, Inc.).

Carr, S.A., Huddleston, M.J., and Annan, R.S. (1996). Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. Anal. Biochem. *239*, 180–192.

Chen, H., Charlat, O., Tartaglia, L.A., Woolf, E.A., Weng, X., Ellis, S.J., Lakey, N.D., Culpepper, J., Moore, K.J., Breitbart, R.E., et al. (1996). Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in *db/db* mice. Cell *84*, 491–495.

Eipper, B.A., Milgram, S.L., Husten, E.J., Yun, H.Y., and Mains, R.E. (1993). Peptidylglycine alpha-amidating monooxygenase: a multifunctional protein with catalytic, processing, and routing domains. Protein Sci. *2*, 489–497.

Erickson, J.C., Clegg, K.E., and Palmiter, R.D. (1996a). Sensitivity to leptin and susceptibility to seizures of mice lacking neuropeptide Y. Nature *381*, 415–418.

Erickson, J.C., Hollopeter, G., and Palmiter, R.D. (1996b). Attenuation of the obesity syndrome of *ob/ob* mice by the loss of neuropeptide Y. Science *274*, 1704–1707.

Fan, W., Boston, B.A., Kesterson, R.A., Hruby, V.J., and Cone, R.D. (1997). Role of melanocortinergic neurons in feeding and the *agouti* obesity syndrome. Nature *385*, 165–168.

Friedman, J.M. (1997). The alphabet of weight control. Nature *385*, 119–120.

Gerald, C., Walker, M.W., Criscione, L., Gustafson, E.L., Batzl-Hartmann, C., Smith, K.E., Vaysse, P., Durkin, M.M., Laz, T.M., Linemeyer, D.L., et al. (1996). A receptor subtype involved in neuropeptide-Y-induced food intake. Nature *382*, 168–171.

Grynkiewicz, G., Poenie, M., and Tsien, R.Y. (1985). A new generation of Ca^{2+} indicators with greatly improved fluorescence properties. J. Biol. Chem. *260*, 3440–3450.

Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillet, D., Muselet, D., Prud'Homme, J.F., Dib, C., Auffray, C., et al. (1996). A radiation hybrid map of the human genome. Hum. Mol. Genet. *5*, 339–346.

Hardman, J.G., Gilman, A.G., and Limbird, L.E. (1996). The Pharmacological Basis of Therapeutics, 9th Ed. (New York: McGraw-Hill).

Hepler, J.R., Kozasa, T., and Gilman, A.G. (1994). Purification of recombinant Gq alpha, G11 alpha, and G16 alpha from Sf9 cells. Meth. Enzymol. *237*, 191–212.

Hudson, T., Stein, L., Gerety, S., Ma, J., Castle, A., Silva, J., Slonim, D., Baptista, R., Kruglyak, L., Xu, S., et al. (1995). An STS-based map of the human genome. Science *270*, 1945–1954.

Huszar, D., Lynch, C.A., Fairchild-Huntress, V., Dunmore, J.H., Fang, Q., Berkemeier, L.R., Gu, W., Kesterson, R.A., Boston, B.A., Cone, R.D., et al. (1997). Targeted disruption of the melanocortin-4 receptor results in obesity in mice. Cell *88*, 131–141.

Ikeda, M., and Matsushita, A. (1980). Reflectance of rat brain structures mapped by an optical fiber technique. J. Neurosci. Meth. *2*, 9–17.

Jacobowitz, D.M., and O'Donohue, T.L. (1978). α-melanocyte stimulating hormone: immunohistochemical identification and mapping in neurons of rat brain. Proc. Natl. Acad. Sci. USA *75*, 6300–6304.

Mountjoy, K.G., Mortrud, M.T., Low, M.J., Simerly, R.B., and Cone, R.D. (1994). Localization of the melanocortin-4 receptor (MC4-R) in neuroendocrine and autonomic control circuits in the brain. Mol. Endocrinol. *8*, 1298–1308.

Munson, P.J., and Rodbard, D. (1980). LIGAND: A versatile computerized approach for characterization of ligand-binding systems. Anal. Biochem. *107*, 220–239.

Ohki-Hamazaki, H., Watase, K., Yamamoto, K., Orura, H., Yamano, M., Yamada, K., Maeno, H., Imaki, J., Kikuyama, S., Wada, E., and Wada, K. (1997). Mice lacking bombesin receptor subtype-3 develop metabolic defects and obesity. Nature *390*, 165–169.

Ollmann, M.M., Wilson, B.D., Yang, Y.-K., Kerns, J.A., Chen, Y., Gantz, I., and Barsh, G.S. (1997). Antagonism of central melanocortin receptors in vitro and in vivo by Agouti-related protein. Science *278*, 135–138.

Oomura, Y. (1980). Input-output organization in the hypothalamus relating to food intake behavior. In Handbook of the Hypothalamus, Volume 2: Physiology of the Hypothalamus, P.J. Morgane and J. Panksepp, eds. (New York: Marcel Dekker), pp. 557-620.

Oomura, Y., Ooyama, H., Sugimori, M., Nakamura, T., and Yamada, Y. (1974). Glucose inhibition of the glucose-sensitive neurone in the rat lateral hypothalamus. Nature *247*, 284–286.

Paxinos, G., and Watson, C. (1986). The Rat Brain in Stereotaxic Coordinates (San Diego: Academic Press).

Qu, D., Ludwig, D.S., Gammeltoft, S., Piper, M., Pelleymounter, M.A.,

Cullen, M.J., Mathes, W.F., Przypek, J., Kanarek, R., and Maratos-Flier, E. (1996). A role for melanin-concentrating hormone in the central regulation of feeding behaviour. Nature *380*, 243–247.

Rosenbaum, M., Leibel, R.L., and Hirsch, J. (1997). Obesity. N. Engl. J. Med. *337*, 396–407.

Rouille, Y., Dugay, S.J., Lund, K., Furuta, M., Gong, Q., Lipkind, G., Oliva, A.A., Jr., Chan, S.J., and Steiner, D.F. (1995). Proteolytic processing mechanisms in the biosynthesis of neuroendocrine peptides: the subtilisin-like proprotein convertases. Front. Neuroendocrinol. *16*, 322–361.

Shughrue, P.J., Lane, M.V., and Merchenthaler, I. (1996). Glucagon-like peptide-1 receptor (GLP1-R) mRNA in the rat hypothalamus. Endocrinology *137*, 5159–5162.

Shutter, J.R., Graham, M., Kinsey, A.C., Scully, S., Lüthy, R., and Stark, K.L. (1997). Hypothalamic expression of ART, a novel gene related to *agouti*, is up-regulated in *obese* and *diabetic* mutant mice. Genes Dev. *11*, 593–602.

Stadel, J.M., Wilson, S., and Bergsma, D. (1997). Orphan G protein–coupled receptors: a neglected opportunity for pioneer drug discovery. Trends Pharmacol. Sci. *18*, 430–437.

Stephens, T.W., Basinski, M., Bristow, P.K., Bue-Valleskey, J.M., Burgett, S.G., Craft, L., Hale, J., Hoffmann, J., Hsiung, H.M., Kriauciunas, A., et al. (1995). The role of neuropeptide Y in the antiobesity action of the *obese* gene product. Nature *377*, 530–532.

Takigawa, M., Sakurai, T., Kasuya, Y., Abe, Y., Masaski, T., and Goto, K. (1995). Molecular identification of guanine-nucleotide-binding regulatory proteins which couple to endothelin receptors. Eur. J. Biochem. *228*, 102–108.

Tartaglia, L.A., Dembski, M., Weng, X., Deng, N., Culpepper, J., Devos, R., Richards, G.J., Campfield, L.A., Clark, F.T., Deeds, J., et al. (1995). Identification and expression cloning of a leptin receptor, OB-R. Cell *83*, 1263–1271.

Turton, M.D., O'Shea, D., Gunn, I., Beak, S.A., Edwards, C.M.B., Meeran, K., Choi, S.J., Taylor, G.M., Heath, M.M., Lambert, P.D., et al. (1996). A role for glucagon-like peptide-1 in the central regulation of feeding. Nature *379*, 69–72.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. *14*, 4683–4690.

Warembourg, M., and Jolivet, A. (1993). Immunocytochemical localization of progesterone receptors in galanin neurons in the guinea pig hypothalamus. J. Neuroendocrinol. *5*, 487–491.

Wijker, M., Wszolek, Z.K., Wolters, E.C.H., Rooimans, M.A., Pals, G., Pfeiffer, R.F., Lynch, T., Rodnitzky, R.L., Wilhelmsen, K.C., and Arwert, F. (1996). Localization of the gene for rapidly progressive autosomal dominant parkinsonism and dementia with pallidoponto-nigral degeneration to chromosome 17q21. Hum. Mol. Genet. *5*, 151–154.

Wilhelmsen, K.C. (1997). Frontotemporal dementia is on the MAPtau. Ann. Neurol. *41*, 139–140.

Wilhelmsen, K.C., Lynch, T., Pavlou, E., Higgins, M., and Hygaard, T.G. (1994). Localization of disinhibition-dementia-parkinsonism-amyotrophy complex to 17q21–22. Am. J. Hum. Genet. *55*, 1159–1165.

Wilm, M., and Mann, M. (1996). Analytical properties of the nanoelectrospray ion source. Anal. Chem. *68*, 1–8.

Wolf, G. (1997). Neuropeptides responding to leptin. Nutr. Rev. *55*, 85–88.

Xu, D., Emoto, N., Giaid, A., Slaughter, C.A., Kaw, S., de Wit, D., and Yanagisawa, M. (1994). ECE-1: A membrane-bound metalloprotease that catalyzes the proteolytic activation of big endothelin-1. Cell *78*, 473–485.

Yen, T.T., Gill, A.M., Frigeri, L.G., Barsh, G.S., and Wolff, G.L. (1994). Obesity, diabetes, and neoplasia in yellow $A^{vy}/-$ mice: ectopic expression of the *agouti* gene. FASEB J. *8*, 479–488.

Zarjevski, N., Cusin, I., Vettor, R., Rohner-Jeanrenaud, F., and Jeanrenaud, B. (1993). Chronic intracerebroventricular neuropeptide-Y administration to normal rats mimics hormonal and metabolic changes of obesity. Endocrinology *133*, 1753–1758.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., and Friedman, J.M. (1994). Positional cloning of the mouse *obese* gene and its human homologue. Nature *372*, 425–432.

GenBank Accession Numbers

The following full-length nucleotide sequences have been submitted to GenBank (accession number in parentheses): *prepro-orexin* cDNAs, human (AF041240), rat (AF041241), and mouse (AF041242); *OX_1R* cDNAs, human (AF041243) and rat (AF041244); and *OX_2R* cDNAs, human (AF041245) and rat (AF041246).